DOCUMENT RESUME

ED 460 953 TM 028 896

AUTHOR	Neill, Monty
TITLE	Testing Our Children: A Report Card on State Assessment Systems.
INSTITUTION	National Center for Fair and Open Testing (FairTest), Cambridge, MA.
SPONS AGENCY	Joyce Foundation, Chicago, IL.; Ford Foundation, New York, NY.
PUB DATE	1997-09-00
NOTE	242p.; FairTest Staff contributors were Laura Barrett, Phyllis Bursh, Mark Earley, Charles Rooney, Robert Schaeffer, Leslye Sneider, and Marilyn Yohe. This full report contains the Executive Summary. For the Executive Summary, independent of the full report, see TM 028 897.
AVAILABLE FROM	National Center for Fair and Open Testing (FairTest), 342 Broadway, Cambridge, MA 02139 ($30). Tel: 617-864-4810; Fax: 617-497-2224; Web site: http://www.FairTest.org.
PUB TYPE	Reports - Evaluative (142) -- Tests/Questionnaires (160)
EDRS PRICE	MF01/PC10 Plus Postage.
DESCRIPTORS	Achievement Tests; Decision Making; *Educational Assessment; Educational Change; Elementary Secondary Education; Objective Tests; Report Cards; Standardized Tests; *Standards; *State Programs; Surveys; Test Bias; *Test Use; Testing Problems; *Testing Programs
IDENTIFIERS	*FairTest

ABSTRACT

In this study, FairTest evaluated how well state assessment practices live up to the promise of high standards without standardization. The practices of states were measured against standards derived from the "Principles and Indicators for Student Assessment Systems," a 1995 publication of education and civil rights groups working through the National Forum on Assessment. FairTest used surveys, interviews, and various documents to evaluate the states and developed a scoring guide to evaluate each state. Survey responses were received from 44 states, and FairTest drew on other documents to evaluate the other 6 states. It found that, after nearly a decade of intensive discussions about the role and nature of assessment, and despite some important improvements, the fundamental approach of state testing has not changed. Labels have sometimes been revised to "assessment," but most state programs still rely on traditional, multiple-choice tests, and most states still use them inappropriately to make high-stakes decisions. Two-thirds of state student assessment systems do not even reach the middle level of system quality. One-third of systems need a complete overhaul, and another third need major improvements. In two-thirds of the states it may be said that testing systems often impede, rather than enhance, genuine education reform. Many states do not base their assessments on their content standards, and too many states use norm-referenced tests rather than tests that compare achievement to state standards. A summary is included for each state. Eight appendixes provide additional information about the surveys, study methodology, and the principles considered, as well as a glossary, a list of abbreviations, a bibliography, and an order form. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 460 953

TESTING OUR CHILDREN

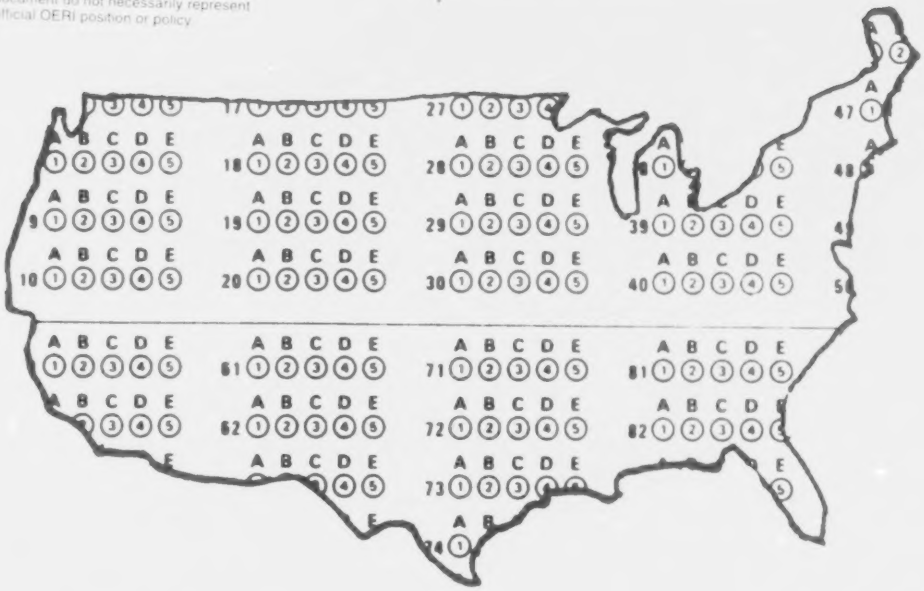
A Report Card on State Assessment Systems

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

C. Schuman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1



by Monty Neill and the Staff of FairTest

FairTest: The National Center for Fair & Open Testing

TM028896

BEST COPY AVAILABLE

TESTING OUR CHILDREN

A Report Card on State Assessment Systems

by Monty Neill, Ed.D.

FairTest Staff who contributed to this report:
Laura Barrett, Phyllis Bursh, Mark Earley, Charles Rooney,
Robert Schaeffer, Leslye Sneider, Marilyn Yohe

September 1997
$30.00

National Center for Fair & Open Testing (FairTest)

342 Broadway
Cambridge, MA 02139
tel: 617/864-4810 fax: 617/497-2224
email: FairTest@aol.com
web page: www.FairTest.org

Testing Our Children: A Report Card on State Assessment Systems
© 1997 National Center for Fair & Open Testing

This report was made possible by grants from the Joyce Foundation and the Ford Foundation. The opinions and conclusions expressed in the report are solely those of FairTest.

Thanks to Steve Ferrara, Ed Reidy and Ramsay Selden, who offered comments at various points. They are not responsible for the conclusions or any errors that may be found in this report. Thank also to the staff of the North Central Regional Education Laboratory and the Council of Chief State School Officers who facilitated our receipt of their data.

Testing Our Children:

A Report Card
on State Assessment Systems

Table of Contents

Introduction 1

Executive Summary 5

- A. Findings in Brief 5
- B. State Performance Levels 7
- C. Patterns and Trends 8
- D. Recommendations 10

State Findings 13

- A. Summary of State Findings 13
- B. Standards 30
- C. Scoring Guide 33
- D. State Data Table 34
  - Alabama 39
  - Alaska 42
  - Arizona 45
  - Arkansas 47
  - California 49
  - Colorado 52
  - Connecticut 55
  - Delaware 57
  - Florida 60
  - Georgia 62
  - Hawaii 65
  - Idaho 68
  - Illinois 71
  - Indiana 74
  - Iowa 76
  - Kansas 78
  - Kentucky 81
  - Louisiana 85
  - Maine 88

Maryland 91
Massachusetts 94
Michigan 97
Minnesota 100
Mississippi 103
Missouri 105
Montana 108
Nebraska 110
Nevada 112
New Hampshire 115
New Jersey 118
New Mexico 121
New York 125
North Carolina 129
North Dakota 132
Ohio 135
Oklahoma 137
Oregon 140
Pennsylvania 143
Rhode Island 146
South Carolina 148
South Dakota 151
Tennessee 153
Texas 157
Utah 160
Vermont 163
Virginia 167
Washington 170
West Virginia 173
Wisconsin 176
Wyoming 179

Appendices 181

A. Abbreviations
B. Glossary
C. Methodology
D. Full Survey
E. Short Survey
F. Excerpts from *Principles and Indicators*
G. Bibliography
H. FairTest Order Form

Testing Our Children:
A Report Card on State Assessment Systems

Introduction

Standardized tests first rose to prominence in the 1920s, the era in which the "factory model" of education established clear dominance. They reinforced that mode of schooling, in which only a few children received a high-quality education, and they were used to sort students hierarchically within that model. The promise of school reform in the 1990s has been to break with that inadequate, often harmful model of schooling. As one part of reaching that goal, assessment must be fundamentally restructured to support high standards without standardization.

In this study, FairTest evaluates how well state assessment practices live up to this promise. We have measured these practices against standards derived from the *Principles and Indicators for Student Assessment Systems,* a 1995 publication by a coalition of education and civil rights groups working together through the National Forum on Assessment.

In broad terms, the *Principles* calls for assessments that are:
- grounded in solid knowledge of how students learn;
- connected to clear statements of what is important for students to learn;
- flexible enough to meet the needs of a diverse student body; and
- able to provide students with the opportunity to actively produce work and demonstrate their learning.

What we have found is that despite nearly a decade of intensive discussions about the role and nature of assessment, and despite some important improvements, the fundamental approach of state testing programs has not changed. Though the labels have often been revised to "assessment," most state programs still predominantly rely on traditional, multiple-choice tests, and many states use them inappropriately to make high-stakes decisions.

Based on a detailed survey and other data sources, we conclude that two-thirds of state K-12 student assessment systems do not reach even the middle level of system quality. One-third of the systems need a complete overhaul, and another third need major improvements if they are to provide support for high quality teaching and learning. The remaining third all have positive components, but still need some improvements.

In two-thirds of the states, then, testing systems often impede, rather than enhance, genuine education reform:

- Rather than holding schools accountable for providing a rich, deep education and reporting on such achievement to the public, most state testing programs provide information on a too-limited range of student learning in each important subject area.

- Rather than supporting and assessing complex and critical thinking and the ability to use knowledge in real-world situations, most state tests continue to focus too much on measuring rote learning.
- Rather than making decisions about students based on multiple sources of evidence, too many states use a single test as a mandatory hurdle.

Since state tests powerfully affect curriculum and instruction, most state testing programs present obstacles to developing high-quality classroom practices and fail to support strong school reform. Some improvements can be seen in the use of writing samples (though these are often themselves narrow) and constructed-response items (though their use remains too limited), and in more attention to bias reduction. However, in most states, these modest changes amount to tinkering at the edges of reform.

In fact, the recent tendency has been to intensify the traditional mode of testing, with higher cut-off scores and more "difficult" exams, without changing the underlying approach. In most state tests, "difficult" means testing student achievement in conventional academic subjects at an earlier age, such as algebra in grade 8. The problem with this approach is not that algebra now may need to be taught in grade 8, but that the kind of algebra tested remains predominantly the memorization of rules and procedures and very limited applications. This approach fails to meet the essence of the math standards of the National Council of Teachers of Mathematics. A similar, flawed approach can be found for every subject.

The negative consequences of relying on traditional tests and using them to control school reform often seem to be the result of continued confusion over the limitations of large-scale assessments. Unfortunately, states often fail to recognize these limitations and expect their tests to be useful in ways they cannot.

Large-scale testing programs are generally not useful in improving a student's immediate learning process, though clearly that is what most parents hope for from assessment. As diagnostic tools, most large-scale tests are blunt, imprecise, and often useless -- but most states claim that diagnosis is a reason for their tests. Because most state tests do not provide any opportunity for sustained and engaged thinking, they are poor tools for shaping or improving curriculum and instruction -- a goal most states claim for their tests. While these exams can provide some information to the public about what students have learned, most do not provide information about whether students can use in their lives the things they have supposedly learned. They thus provide limited accountability information.

Despite these extreme limitations of state testing programs, the cumulative effect of the multiple uses of these tests is that the exams largely define the purpose and processes of schooling in most states. They affect not only curriculum and instruction, but also the culture of learning, student motivation, and the underlying conceptions of what learning is and how humans learn. Driving school reform with traditional tests will not succeed if the nation really wants all children, not just the children of the wealthy, to gain an education that challenges their minds and spirits, that assumes not only that they can learn some skills but can learn to use their learning as active participants in a democratic society.

There is an alternative. The *Principles and Indicators* calls for large-scale assessments that combine sampling from classroom-based assessment data, such as portfolios and learning records, with performance exams administered to samples of students. In this way, essential standards are promoted and accountability information is gathered, while schools are encouraged to become communities of learning that support all their students. Only one state, Vermont, approaches this model, though elements of the assessments in a few other states are headed in this direction.

Fundamental assessment reform is still feasible. What is lacking is not the technical know-how, though much remains to be learned in that domain, but the political will. The responsibility for improving assessment programs rests first of all with policymakers -- governors, legislators, boards of education. It rests secondly with all those who can educate, or influence, the policymakers -- educators, parents, community and business leaders, testing experts, state education staff, and the voting public. That makes achieving real assessment reform an education and organizing project. Only with an informed and active community, as well as educated policymakers, can deep reform be created and sustained, including the necessary transformation of state assessment programs.

Executive Summary:
State assessment systems in light of the *Principles and Indicators for Student Assessment Systems*

Across the nation, state testing systems powerfully affect curriculum, instruction, school cultures, and the quality of education delivered to our nation's children. They can either support important learning or undermine it.

This study evaluates how well state assessment systems support and help improve student learning. FairTest based its evaluation on standards derived from the *Principles and Indicators for Student Assessment Systems*. This document was developed by the National Forum on Assessment to help guide assessment reform and has been signed by over 80 education and civil rights groups (see Appendix F). To gather data, FairTest used surveys, follow-up interviews, and various documents (see Appendices D and E).

A. Findings in Brief

Among the findings of this study are the following:

1) On a five-point scale for scoring state assessment systems, two-thirds of state K-12 student assessment systems do not reach even the middle level of system quality: one-third of the systems need a complete overhaul and another third need major improvements if they are to provide support for high quality teaching and learning. A few states have made good progress, reaching level 4, but only one, Vermont, has reached the top level.

2) While most states now have content standards, many state tests are not based on their standards, and many important areas in their standards are not assessed.

3) Most states rely far too heavily on multiple-choice testing and fail to provide an adequate range of methods for students to demonstrate their learning. This results in not assessing important areas and creating the likelihood that those areas will not be taught.

4) Too many states use norm-referenced tests (NRTs), which compare students to reference groups and not to achievement on state standards. These tests fail to assess important areas of the standards and encourage grouping and instructional practices that historically have failed to provide many students with a strong education.

5) The state testing burden is often too heavy, with students repeatedly tested in the same subjects. A few states test students in almost every grade. For accountability purposes, extensive testing is not necessary.

6) Seventeen states use a single test as a necessary requirement for high school graduation, violating the AERA/APA/NCME standards for good assessment practice, ensuring unfair treatment of many students, and increasing the likelihood that narrow tests will dictate curriculum and instruction. Districts may use state tests as graduation or grade promotion hurdles. An additional five states currently plan to implement such tests, two of which plan to allow an alternative option.

7) Most writing assessments require students to respond to a single prompt, fostering and reporting a limited conception of writing. Writing must serve many purposes and therefore take many styles. A major problem here is the potential reduction of writing instruction to fit the state exam.

8) Rich assessment techniques, such as portfolios and performance events, are rarely used by states. Thus, important areas of learning are not assessed and important signals are not sent to schools about what students should be learning and how assessment can support that learning.

9) Very few states use sampling for accountability, public reporting, and program improvement purposes, even though it provides accurate data, is less expensive and less intrusive, and allows greater use of portfolios and performance events.

10) Most states use tests for student diagnosis and for improving curriculum and instruction, even though most large-scale tests are crude tools for diagnosis and too narrow to support high quality curriculum and instruction.

11) A solid majority of states have bias review panels, often with significant authority to delete or revise items on state-made tests, but some do not. This is a positive development.

12) States tend not to adequately assess or include in state reports students with Individual Education Plans (IEP, e.g., "special education") and students with Limited English Proficiency (LEP). Inclusion of all categories of students, using appropriate assessments, is necessary for proper program evaluation and ensuring proper education for these students. The recently reauthorized federal Individuals with Disabilities Education Act will require all students with disabilities to be assessed appropriately, but such provisions do not exist for LEP students.

13) States are generally quite weak in providing adequate professional development in all aspects of assessment to teachers and other educators. Such teacher education, particularly in classroom assessment, is fundamental to assessment and broader school reform.

14) Few states evaluate teacher competence in assessment or study district, school and classroom assessment practices or their impacts. Thus, they lack information to help improve the quality of assessment at all levels and to halt harmful practices.

15) Student and parent rights, such as the ability to review tests after completion, to challenge flawed items or to appeal scores, exist unevenly. Such rights are fair in themselves and also

help parents better understand assessment and education in general and to view themselves as important partners in their children's education.

16) Reporting to the public and educating the public about assessment are often limited, and few states report in languages other than English, even if they have a large number of residents who do not speak or read English.

17) State reviews of their assessment systems need substantial improvement. Most do not study the impact of testing on curriculum, instruction, or graduation rates; and most do not review whether their assessments measure the ability of students to think critically or in complex ways in the various subject areas. In an era in which testing is proposed as a fundamental tool for school reform, states often cannot even be sure whether increasing scores are based on real learning gains or teaching to the test.

B. State Performance Levels

Using a scoring guide, FairTest evaluated each state. The list below reports which states scored at each level of the scoring guide. The scoring guide is found in the section on state findings, and details for each state are provided in the full report.

How the States Scored

Level 5	**A model system.** Vermont
Level 4	**State assessment system needs modest improvement.** Colorado, Connecticut, Kentucky, Maine, Missouri, New Hampshire
Level 3	**State assessment system needs some significant improvements.** Illinois, Kansas, Maryland, Michigan, Oregon, Pennsylvania, Rhode Island
Level 2	**State assessment system needs many major improvements.** Arkansas, California, Idaho, Indiana, Massachusetts, Minnesota, Montana, Nebraska, Nevada, North Dakota, New Jersey, New York, Ohio, Oklahoma, South Dakota, Texas, Washington, Wisconsin
Level 1	**State assessment system needs a complete overhaul.** Alabama, Alaska, Arizona, Florida, Georgia, Hawaii, Louisiana, Mississippi, New Mexico, North Carolina, South Carolina, Tennessee, Utah, Virginia, West Virginia
	Not scorable. Delaware, Iowa, Wyoming

BEST COPY AVAILABLE

C. Patterns and Trends

A few basic patterns and trends over the past decade, based on a comparison between this and other reports, can be discerned. These include:

1) The amount of testing done by the states appears not to have changed very much, though it seems to vary year to year as states alter their testing programs.

In its 1988 report, *Fallout from the Testing Explosion*, FairTest found, by comparing the numbers of tests administered to school enrollments, that states were administering .42 tests (which may include more than one subject area) per year per student. (District testing, primarily achievement and special needs testing, raised the average to about 2.5 tests per student per year.)

To identify current testing frequency, we examined CCSSO/NCREL data for various grades tested over the past few years. The 1993-94 data show that the states tested a total of 278 grades, or an average of 5.56 grades. (This assumes a state uses only one test at a grade level, but some do use more than one test at a given grade level). With 13 grades, this averages to .43 tests per year per student. In 1994-95, the numbers declined to 243 grades tested, or an average of 4.86 grades or .37 tests per year per student. But in 1995-96, the numbers were back up slightly, to 264 grades tested, or 5.28 tested grades per state and .41 tests per student.

As the means of determining the amount of testing was different in *Fallout*, the numbers are not directly comparable, but they give a rough sense of the stability of the amount of state testing over time.

2) *Fallout* reported that 11 southern states (Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee and Virginia) tested more often than did the rest of the nation. This continues to be true. In 1995-96, those 11 states tested in 7 grades on average. The other states which are part of the Southern Regional Education Board (SREB) actually now test even more: Texas tested at 9 grades, Maryland at 8, Oklahoma at 8, and West Virginia at 11, bringing the SREB average to 7.5 grades, substantially higher than the national average of 5.28 grades. Another way of looking at it is that 30 percent of the states do 43 percent of the testing.

3) These states are also more likely to mandate high school graduation tests. Of the 15 SREB states, 11 have graduation exams. Only six of the 35 states outside the South have such a test.

4) The number of states with high school exit exams declined in the 1990s but is now growing again. In 1989, *Education Week* (May 10) reported 23 states had or intended to have these exams. By 1994-95, CCSSO/NCREL reported that 17 states had mandatory exit exams. FairTest confirmed this number, but also found that five more states plan to adopt such a requirement.

5) Other than southern states, half the states with high school exit exams are in the northeast: New Jersey, New York and Ohio are joined by Hawaii, New Mexico and Nevada. The states that soon will require such tests are Alaska, Arkansas, Delaware, Indiana, and Massachusetts. This will bring the total number of states that have or are planning to have exit exams to 22 -- about where it was at the end of the 1980s.

6) *Fallout* noted that large cities tested more often than smaller cities or rural areas. Combined with the data on southern states, this suggests that areas with large proportions of African Americans are most likely to test heavily. States with relatively large proportions of African Americans are more likely to administer high school exit exams.

7) It also appears that the 15 SREB states, with the notable exception of Kentucky and Maryland, are less likely to use constructed-response or performance assessments (excepting writing to a prompt) than is the nation as a whole. States with mandatory high school exit tests also appear less likely to use constructed-response or performance assessments, again excepting writing to a prompt (see Fairbanks & Roney). These findings may be starting to change as more states use constructed-response items, including in graduation tests.

8) Southern states also are more likely to use NRTs. Thirty-three states use an NRT, including those which sample (North Carolina and Maryland), those which require it of districts (Nebraska) or pay for districts use of one (California and Iowa). All of the 15 SREB states except Texas use an NRT. Roughly half of the remaining states use an NRT (19 of 35).

9) All told, there appears to be a "southern effect" which includes high-stakes testing, a heavy testing load, use of an NRT, and relatively less use of constructed-response and performance assessments. As a group, the southern states still are the nation's poorest region, so this is also a "poverty effect." Results of the National Assessment of Educational Progress continue to show the southern region lagging behind the rest of the nation in terms of measured educational achievement.

Since there is evidence that using performance assessments signals or spurs a shift toward teaching and assessing more challenging, cognitively complex material, then the southern states could be left behind once again. As the negative effects of teaching to narrow tests most powerfully affect schools with large proportions of minority-group and low-income children, such students in these states are particularly at risk of continuing to receive a low-level education that will not prepare them well for their adult lives. Students in large cities that also emphasize teaching to traditional tests face the same risk.

Unfortunately, these southern states, along with others, are caught in a vicious circle. Low scores lead to more tests and higher stakes. More tests and higher stakes lead to more intense "teaching to the test." Teaching to narrow, multiple-choice tests leads to an overemphasis on rote memorization at the expense of higher order thinking skills. In this way, tests themselves are part of the problem, not the solution.

Fortunately, several states across the country are trying to break this cycle. They are increasing their use of assessments that measure genuine knowledge, not simply facts, and that evaluate a student's performance on multi-faceted tasks, not simply his or her ability to select the preferred response from a list of possible answers. They are also paying great attention to professional development so that teachers learn well how to use performance assessments and portfolios in their classrooms. This facilitates a bottom-up approach to school reform rather than relying solely on top-down, test-driven initiatives.

If these alternative assessment systems are allowed to survive the growing pains of their early years, they will provide educators in other states with valuable knowledge about how to alter their assessment systems. Perhaps then most of the states, not just a few, will move beyond tinkering at the margins and will completely overhaul their state assessment systems.

D. Recommendations

These findings establish the framework in which fundamental assessment reform must take place. A great deal has been learned, some of it from pioneering efforts in a few states, some of it in districts, most of it in schools and classrooms. What is lacking is not the technical know-how, though certainly problems remain, but the political and social will to recreate assessment as part of reinventing education.

If large-scale assessments are to support excellence and equity in education, FairTest concludes that underlying conceptions and basic practice in most states need to be fundamentally changed and brought into alignment with the *Principles and Indicators for Student Assessment Systems* as follows:

1) Base all state (or district) assessments of student achievement on clear standards.

2) Employ multiple methods of assessment, limiting multiple-choice to no more than one quarter of test-takers' scores.

3) Rely on methods that allow students to demonstrate understanding by applying knowledge and constructing responses and that ensure assessment of complex and critical thinking in and across subject areas.

4) Do not use norm-referenced tests, or limit their use to very light sampling.

5) Do not make high-stakes decisions, such as high school graduation, using single exams as a hurdle. Rely on multiple sources of information.

6) Employ sampling procedures to collect information on large populations, using performance and portfolio assessments.

7) Rely on sampling from classroom-based work as a key component of large-scale information on student achievement, including work which allows individual choices and expressions of knowledge and provides students the opportunity to evaluate their own work.

8) Enhance efforts to appropriately include all students in assessments and reporting, and report disaggregated data by important population groups.

9) Ensure adequate professional development in assessment, particularly in classroom and performance assessment, for both teachers and students in education schools.

10) Systematically involve teachers and other educators in developing and scoring performance assessments and portfolios.

11) Institute comprehensive reviews and use the results to improve assessments.

State Findings

To evaluate the specific characteristics of state assessment programs, FairTest adapted the *Principles and Indicators* to create standards and indicators appropriate for large-scale assessment. The standards are:

Standard 1: Assessment supports important student learning.
Standard 2: Assessments are fair.
Standard 3: Professional development.
Standard 4: Public education, reporting, and parents' rights.
Standard 5: System review and improvement.

The following explains the basic purpose of each standard and indicator and why it is important, summarizes the findings from across the states, and discusses the implications of each finding. Forty-four states responded to the FairTest survey, providing relatively complete information for the evaluation process. For the remaining six states, FairTest relied on other sources which provided substantially less data and no information at all on many of the indicators in the standards.

A. Summary of State Findings

Standard 1: Assessment supports important student learning.

The *Principles* states: "Assessment systems provide useful information about whether students have reached important learning goals....They employ practices and methods that are consistent with learning goals, curriculum, instruction, and current knowledge of how students learn. No assessment... is used that narrows or distorts the curriculum or instructional practice."

Large-scale assessments should be used to gather data for program improvement and to report program-level data to the public. Most other assessment purposes, such as individual student diagnosis, reporting individual progress and determining who should graduate, are better left to schools and teachers. Large-scale assessments are necessarily blunt instruments, and so should be used sparingly, with caution, and for purposes in which large-scale information makes sense.

Unfortunately, state programs often undermine important student learning through overuse of multiple-choice testing and norm-referenced tests, under-utilization of performance assessments and portfolios, high-stakes uses of single exams, and over-testing. The assessments are often so limited as to undermine content standards (which most states have adopted) by not assessing important areas in the standards. Though one of the most commonly stated purposes of state assessments is "program improvement," most state assessments are not adequate for helping to develop high-quality education programs.

Some states do not have state testing programs. The *Principles* does not recommend either state standards or state assessments and recognizes these can be undertaken at the district level. However, FairTest concludes that states which rely on district testing should then evaluate district practices and support improvements at the district level. In some states without formal state programs, the state mandates district assessments. In these cases, the mandate is effectively a state program and can be evaluated as such. The state also can be evaluated in terms of its direct activities or support for districts on the issues of fairness, professional development, reporting, and evaluation of the assessment program.

1.1. Assessments are based on and aligned with standards. Students deserve to have clear statements of what they are expected to learn and the opportunity to master that material. States should have standards if they have state exams, and the exams should assess comprehensively and in a balanced fashion the content that is in their standards. If a state mandates district achievement testing, it also should mandate that those tests be based on state or district standards.

While most states now have standards and increasingly report that their assessments are aligned to the standards, too often important areas in these standards are not assessed. This is largely because of limited assessment methods, particularly over-reliance on multiple-choice testing. Some states acknowledged this, noting such things as "multiple-choice cannot assess all areas in the standards" or even noting a percentage of the standards that is measured. Others simply claim that their multiple-choice tests are matched to the standards. The reality is that most state tests do not comprehensively and in a balanced manner assess students to high-quality content standards.

The clear dangers are that what is not tested is not taught, that what is tested is the lower levels of the standards, and that curriculum is therefore reduced to its lower levels. Based on previous experience, the curriculum is most likely to be narrowed in schools and districts where students do not perform as well on the tests. The consequence, which has been observed in various research studies, is often to continue to deny a challenging and engaging education to those students who have historically not been well-served by public schooling, particularly students from low-income families and students of color. As discussed in Standard 5, it appears that few states seriously investigate this issue.

1.2. Multiple-choice and very-short-answer (e.g., "gridded-in") items are a limited part of the assessments; and assessments employ multiple methods, including those that allow students to demonstrate understanding by applying knowledge and constructing responses. These requirements are strongly stated in the *Principles*. FairTest recommends that not more than one quarter of a student's score in any subject be obtained from multiple-choice and very-short-answer items.

Serious critical and complex thinking in subjects, real-world problem solving, and application of knowledge cannot be assessed adequately with multiple-choice items. Further, as teachers tend to teach to state exams, focusing instruction on multiple-choice tests limits curriculum and instruction in ways that deny students opportunities to think, tends to narrow the range of instructional practices, and reduces student motivation to learn -- all of which combine to undermine both excellence and equity. Using such tests for "diagnosis," as many states report doing, compounds the problem: they are too limited a measure for useful diagnosis for most instructional purposes.

Most of a score should come from methods that allow students to apply knowledge, solve complex problems, and demonstrate thinking within a subject. Such an approach enables assessment to better match high-quality standards. These are also practices that are more compatible with how humans learn. Additionally, using multiple methods allows students with different learning styles an opportunity to demonstrate their achievement and enables the assessment of content or skills that are not assessed well by other methods.

Unfortunately, most states rely too heavily on multiple-choice items and fail to use a reasonable range of assessment methods. Excluding writing assessments, of the 50 states, 26 rely entirely or nearly entirely on multiple-choice. Another 16-18 rely mostly on multiple-choice (have less than half their scores derived from constructed-response items; in two states, the proportions were not clear but appear to be around the one-half point). Only 6-8 states have less than half multiple-choice items.

Using a variety of methods does not require that multiple-choice be one of them. Rather, the mix could include short and extended constructed-response items, performance events, and portfolios.

Most fundamental is that the actual tasks and items are of high-quality. This study could not evaluate the quality of the items or whether taken together they comprise a high-quality assessment.

Thirty-eight states have writing assessments (including Vermont, where it becomes mandatory next year). However, with rare exceptions, the writing is simply responding to a pre-selected prompt, with students allowed no opportunity even to select from a set of prompts. Only three have portfolio writing assessments. Unfortunately, response to a prompt creates a very narrow picture of writing and encourages teaching geared to an arbitrary formula, such as the five-paragraph "essay." This is also an equity issue, as students who happen to be interested in or knowledgeable about the one particular topic will have an unfair advantage. Instead, more than one form of writing should be assessed and students should have a choice of prompts. An additional issue is the time allowed for response, which in some states is too short. Some research suggests that student performance improves with extended time for response, a point that is relevant not just to writing.

1.3. Assessments designed to rank order, such as norm-referenced tests (NRT), are not used or are not a significant part of the assessment system. These tests are constructed to compare students rather than to see how well students achieve according to standards. Norm-referencing is rooted in the concept of the "bell curve." The use of comparisons and the bell curve, which by definition place half the students "below average" or even "below grade level," suggests that many students will not learn to high levels and meet state standards. The use of NRTs often encourages tracking, sorting and low expectations.

Thirty-three states use NRTs, some as the major state component and some together with a criterion-referenced test (CRT); two of the 33 use them only on a sampling basis. Some NRTs now include, as an option, constructed-response items, but almost all states which use commercial NRTs still use exclusively multiple-choice versions. A few states report their results according to state norms. This is also inappropriate; their exams should be constructed around state standards and be reported in terms of those standards.

1.4. The test burden is not too heavy in any one grade or across the system. Students often are tested far more frequently than is needed to produce data for program improvement or accountability. Consequently, valuable classroom time is wasted preparing for and taking exams that serve no useful purpose. A reasonable system is one in which students are assessed in a subject once at each level (elementary, middle, high), as is now required by the federal Title I program. A model system would rely on sampling.

The test burden required by states varies greatly, from a few tests in a few grades, to many subjects tested in a few grades, to a few subjects tested in many grades, to many subjects tested in many grades. The state test burden is often unnecessarily heavy. Many districts add yet more standardized tests to the state exams, so what appears to be a reasonable burden in some states may be, in most of that state's districts, a high burden. Few states, however, even survey district assessment practices.

FairTest has not addressed the issue of how many subjects should be tested but recommends that if more than two subjects are tested, the burden should be spread over several grades (e.g., English language arts and math in grade 4, science and social studies/history in grade 5). Except for comments in a few state reports, we also did not address the issue of the amount of time devoted to testing.

1.5. High-stakes decisions, such as high school graduation for students or probation for schools, are not made on the basis of any single assessment. The AERA/APA/NCME *Standards for Educational and Psychological Testing* state at Standard 8.12: "[A] decision or characterization that will have a major impact on a test taker should not automatically be made on the basis of a single test score." Similar statements can be found in numerous other test use guidelines, including the *Principles*. FairTest concludes that no single test should act as a barrier to graduation.

By "single assessment" we mean "hurdle" -- as in a track race in which each and every one must be cleared. Thus, using a test as a stand-alone hurdle means it must be passed for graduation or promotion -- even if there are, as is typical, multiple opportunities to clear the hurdle.

However, 17 states use a test as a high school graduation requirement. Two states include state assessments as part of determining grade promotion. Some districts also may use state assessments in determining grade promotion or graduation, though the information on this is largely anecdotal. States sometimes report the tests are also used for placement purposes, which would include tracking and which certainly can be high-stakes uses. States need to monitor districts to ensure tests are not misused in making decisions.

The number of states with graduation exams has been fairly stable at about 17 for a few years. At the turn of decade, FairTest compiled a list of 24 states that had or intended to have such requirements, so by the middle of the decade substantial progress had been made. However, in the past several years, a stronger push has come from a number of quarters to implement graduation exam requirements. It now appears that by about 2000, at least five more states will have such policies in place.

For students, this is substantially a fairness issue. Individuals should be judged on the basis of their accumulated work, not their score on a one-shot test. Similarly a range of information should be considered in evaluating programs. Decisions should not be triggered solely by results on tests. In fact, for most states which have established potentially serious consequences for schools or districts, such as probation or takeover, scores are one of a number of factors which trigger investigations prior to actions, which is as it should be. At a minimum, states with high-stakes tests for individuals should apply this approach.

A second reason for this standard is that the higher the stakes, the more likely the tests will control curriculum and instruction. Graduation tests are usually entirely or almost entirely multiple-choice, sometimes with a writing sample added in, so the issues raised around multiple-choice tests pertain with most force to these high-stakes exams. Any stakes, starting with public reporting and increasing through a variety of sanctions and rewards for schools or students attached wholly or in part to test results, can begin to cause instruction to focus on the content and method of the tests. If this approach to focusing instruction is to be valid, then the exams must adequately assess the range of knowledge, understanding, skills and abilities that schools seek to teach. In addition, the tests should change every year to prevent narrow teaching to one set of items. Few state exams meet these requirements.

1.6. Sampling is employed to gather program information. Sampling, rather than testing every student with an entire exam, is a reasonable solution to a fundamental quandary in large-scale assessment: how to use time-consuming and expensive performance events and portfolios as a major source of data, given limited funds. Matrix sampling, in which an assessment is divided into parts and each test-taker is administered only one of the parts, can be particularly efficient for exams.

Only a few states make even limited use of sampling. Missouri is probably dropping sampling from its new system, Maine uses sampling in some subjects but may be switching to testing every student, and North Carolina and Maryland use sampling with an NRT. The best case is Vermont, which re-scores samples of student portfolios (in which every student has a portfolio) to obtain state-level data. However, because it has many small schools, Vermont will not use sampling in its new performance exams, but will test every student.

The essential problem, however, is political -- the perception that parents and the public want every child tested and scored. So long as this remains the policy imperative, it is unlikely that much progress will be made in using instructionally appropriate assessment methods. That is, choosing to test every child inexpensively requires the use of narrow testing methods. This educational cost is generally not explained to the public so as to create an informed discussion of the trade-offs.

The educationally superior alternative is to use large-scale assessments employing statistically sound samples to report program data and to have individual data gathered and reported by schools. Schools also would make high-stakes decisions and certify student achievement, such as for high school graduation.

1.7. The evaluation of work done over time, e.g., portfolios, is a major component of accountability and public reporting data. As emphasized in the *Principles*, students should be evaluated primarily on the basis of their regular classroom work, accumulated over time, rather than on the basis of one-time tests. This enables examination of much richer information than can be obtained from "snap-shot" tests. It also supports fairness by allowing and encouraging a greater variety of student work.

Only six states use portfolios at all as part of the state testing program, though a number of other states are supporting districts and schools in developing portfolios. One obstacle has been the complexity of gathering an appropriate selection of a student's work and evaluating it reliably. The education of scorers to respect diversity while insisting on quality also is essential. Nonetheless, the major obstacle appears to be the political decision that the state should assess each individual student, rather than to sample, thus making use of portfolio assessment for program evaluation and accountability very expensive.

1.8. Students are provided an opportunity to comment on or evaluate the instruction they receive and their own learning. Principle 1 notes that self-reflection is an important element of assessment and learning and should be part of the assessment system. While this is primarily a classroom issue, it has a place in large-scale assessments, for two reasons. First, its inclusion signals that self-reflection is important. Second, the information received can be used in evaluating what works and why in curriculum and instruction.

Only a few states include this option, usually in a survey attached to the state exam. Similarly, only a few states survey teachers or administrators about instruction and assessment (see Standard 3).

1.9. Appropriate contextual information is gathered and reported with assessment data. Such data includes information about the actual curriculum and instruction provided to students, the instructional and physical resources, demographic data, information on spending and the teaching force, class size, student mobility, tracking and placement policies, and other outcome information.

It appears that few if any states gather much of this important information. Not one state indicated it gathered or reported such contextual data. It is possible that the information is gathered elsewhere within state education departments, but it is likely that much of the desired information is not obtained or is not used in conjunction with assessment data.

Collecting contextual information is called for in the *Principles* because the information can be used in program evaluation, such as when interpreting achievement data. Additionally, while it would be inappropriate to justify low scores by reference to demographics, serious efforts at school reform require providing every student with an adequate and appropriate opportunity to learn. Thus, gathering contextual information is essential for using assessment results to improve programs rather than to simply report, praise or blame.

Standard 2: Assessments are fair.

Assessment systems must not limit students' present or future opportunities and must provide all students with a reasonable and fair opportunity to demonstrate their achievement. The *Principles* states: "Assessments are fair when every student has received equitable and adequate schooling, including culturally sensitive curriculum, instruction and assessment that encourage and support each student's learning....Assessment results accurately reflect a student's actual knowledge, understanding and achievement. Assessments are designed to minimize the impact of biases."

In some regards, states have made progress, particularly through bias and sensitivity review panels that often have the power to delete or revise items. Increasingly, states are aware of the need to provide adequate assessments to students with exceptional needs, but actual progress on such assessments has been limited. For students with Individual Education Plans (IEPs), this should soon change under the impetus of the recently revised Individuals with Disabilities Education Act (IDEA) federal legislation. The fairness standard also says that states do not make important decisions based on a single test score and that they provide students with opportunities to be assessed with multiple methods. On these issues, states are not making much progress.

2.1. States have implemented comprehensive bias review procedures. Bias in assessment renders an assessment invalid for the population against whom the assessment is biased. This is true not only because biased items fail to accurately measure all students' learning on that item, but also because biases can undermine how a student responds to an entire exam. Bias can include race, gender, socioeconomic class, culture, language, rural/urban, handicapping status, and sexual orientation. To guard against bias, committees -- with the authority to remove or modify items -- should examine individual items and the exam as a whole. Statistical procedures that can help detect biased items should also be used.

Most states have a bias review procedure. Bias reviews typically consider race and gender; some states reported considering disability or linguistic and cultural background; only a few states report considering other issues, such as socio-economic status.

Most have a separate bias review committee, though sometimes a content committee will examine items and the whole assessment for bias. For commercially published tests, states usually rely on bias review by the test maker, which often includes both committees (with unknown authority) and statistical studies. Thirteen of the states responding to the full FairTest survey reported doing statistical analyses of tests for bias, which should and sometimes does include studying tests both before and after administration.

In general, state and commercial exams appear to do fairly well in terms of identifying overtly biased items. Broader issues, such as the kinds of content in the composition of the test and the possible impact of the presence or absence of certain content (even if not overtly biased) on test takers, is studied in some states, but not in others (on this, we did not obtain much information).

2.2. Assessment results should be reported both for all students together and with disaggregated data for sub-populations. Failure to include all students in reports sends the message that they are less important and need not be considered. But it is also important to report disaggregated data in order to track the progress of groups which historically have not been well served by school systems.

A majority of states do some reporting of data disaggregated by demographic categories. States most commonly report by race and gender, while a few report socio-economic class. As noted below, states vary greatly in their reporting of students with IEPs or with limited English proficiency (LEP). In general, states need to do more to present disaggregated data, including at the district and school levels.

2.3. Adequate and appropriate accommodations and adaptations are provided for students with Individual Education Plans (IEP).

2.4. Adequate and appropriate accommodations and adaptations, including translations or developing assessments in languages other than English, are available for students with limited English proficiency (LEP).

States have only recently begun to consider including all students in their assessments. According to the National Center on Educational Outcomes (NCEO), many states still do not know how many students with IEPs are or are not assessed. Many states assess only a small percentage of their IEP students. The situation is often worse for students with LEP.

The 1997 reauthorization of the Individuals with Disabilities in Education Act (IDEA) requires states to develop standards for students with special needs that are coordinated with any state standards for all children; and to include students with IEPs in their accountability systems, including assessments, with appropriate accommodations and, if necessary, alternate assessments. They are to be reported both in general reports and disaggregated. The new legislation therefore will bring the states closer in line with this standard. It is less certain that similar progress will be made in assessing students with LEP, as they are not included in the legislation.

A critical issue will be whether the assessments will be appropriate for the students. Not all students can reasonably be assessed with regular assessments. Some students require accommodations to make the results fair and meaningful. Still others may require alternate assessments. However, whether, or the extent to which, accommodations may alter the meaning of the assessment is not fully understood, and research is being done on this issue. Nonetheless, fairness requires that students with an IEP or who are LEP be assessed in terms of state standards and with appropriate assessments. The results should be included in regular reports wherever possible, as well as reported separately, so the success of programs for students with special needs can be evaluated. Requiring all students to be assessed and included in regular reports can also lessen the tendency to place some students in special programs so that they will not be assessed, enabling school or district scores to appear higher.

While states show a great range on this category, in general they do not yet properly include and assess IEP and LEP students.

FairTest attempted to obtain data on the percentage of students in each state with an IEP or who are LEP. The intent was to compare this with the percentage tested with LEP/IEP. However, too few states reported the first part for us to know for most states what percentages of students with IEP or LEP are not assessed. According to a recent NCEO report, many states do not know how many students are excluded. However, from the data available, it appears that large numbers of IEP and LEP students are not included in assessments in most states.

The accommodations or modifications available also vary greatly. The fewest tend to be available on commercial NRTs. Alternative assessments, such as the portfolio option used for more severely disabled students in Kentucky, are also very rare. Kentucky is the only state to assess all students with IEPs; no state assesses all students with LEP.

Though always desirable, assessments in languages other than English are particularly to be expected in states with high proportions or numbers of LEP students. California, Texas, New York, Florida, Illinois, Arizona, New Mexico, New Jersey, Michigan and Massachusetts have more than 40,000 students with LEP, and Washington and Oklahoma have over 25,000 LEP students. (See reports from George Washington University Evaluation Assistance Center East.) Only a few of these states provide assessments in languages other than English.

States vary in their reporting procedures for students with LEP and IEP. Some include them in regular reports, some publish separate reports, some do both, and some do neither. FairTest supports the approach of inclusion in regular reports and disaggregated reporting.

Finally, students with special needs should be included in the population for whom assessments are designed and in the population on whom tests are tried out. A few states reported doing this, though this question was not specifically asked. Additionally, professionals with knowledge of disability and language issues should be involved in developing the assessments.

2.5. Multiple methods of assessment are provided to students to meet needs based on different learning styles and cultural backgrounds. Students have varying learning styles and ways of expressing their knowledge and abilities. Different cultures reinforce different ways of organizing and demonstrating knowledge. Assessment should respond to these issues, as is recognized also in the *Standards for Educational and Psychological Measurement.*

Only a handful of states reported that they considered different learning styles or cultural variations, usually states that had included constructed-response items. It is likely that large-scale assessments, particularly exams, can only address this issue in a limited fashion. Even if a variety of methods are used in one exam, students can still be penalized for not doing well in one format compared with others. However, having multiple methods on an assessment at least conveys the need to use different methods in the classroom and provides some opportunities for students to use different modes of presenting knowledge.

2.6. Students are provided an adequate opportunity to learn about the assessment. Knowing about the format as well as the content of an assessment can be important to doing well. Knowledge about test methods should not be a source of score differences on measures of achievement. Thus, all students should be equally well prepared to use any methods employed on a large-scale assessment, and states should ensure that students are informed and prepared.

Most states make an effort to provide information to students, but the extent and quality of the information appears to vary greatly. As new assessment methods come into use, it is particularly important for states to ensure that students understand how to respond to those methods. Though states with new methods often provide examples for teachers to use with students, it is not clear whether these efforts actually ensure equity in format preparation among students.

Note: It is important to have a strong representation in the assessment development process of people from minority groups which will be assessed. Preferably, they would be over-represented in committees that design assessments and write and evaluate items, so that they can attain a critical mass to influence test construction. The survey did not address this issue.

Standard 3: Professional development.

The *Principles* explains, "Assessment systems depend on educators who understand the full range of assessment purposes, use appropriately a variety of suitable methods, work collaboratively, and engage in ongoing professional development to improve their capability as assessors."

States should ensure that incoming teachers have been adequately prepared to assess their students and that currently practicing teachers are competent assessors. States should provide or ensure that districts provide continuing professional development to meet this goal. Professional development is often enhanced by teachers' participation in developing and scoring performance tasks, so states should consider this value when they consider whether to contract out scoring.

The states are generally quite weak in providing adequate professional development in all aspects of assessment to teachers.

3.1. States have requirements for beginning teachers and administrators to be knowledgeable about assessment, including appropriate classroom practices. Without such requirements, schools of education may not require such preparation, leaving incoming teachers unable to adequately assess their students.

Most states have no assessment knowledge requirements for incoming teachers, and in particular they have no requirements for them to become competent in performance and classroom assessment. Licensing exams may have a few questions about assessment, but this is not a sufficient basis for assuming competence.

3.2. States provide sufficient professional development in assessment, including in classroom assessment. The state should ensure that teachers receive sufficient professional development in assessment. This support should be extensive and systematic. If states delegate this to districts, they should facilitate districts' ability to provide necessary professional development.

While most states provide some sort of professional development, most of it is neither extensive nor systematic. Various studies have suggested that even the best states find their efforts insufficient to meet demand when major reforms in standards or assessments occur. Since strengthened classroom assessment capabilities and restructured large-scale assessments are called for in the *Principles*, states need to do a great deal more to provide professional development and the opportunity for professional collaboration.

3.3. States survey educators about their professional development needs in assessment and evaluate their competence in assessment. These are means to determine what professional development is most needed. The evaluations should be done on an occasional and sampling basis to determine whether the professional development has succeeded and teachers are able to use assessments to support and evaluate student learning.

States rarely ask educators what they need regarding professional development in assessment, nor do they evaluate teacher competence in assessment. A few states have started to address this gap by surveying at least a sample of teachers about their needs and their practices as part of the state assessment program.

3.4. Teachers and other educators are involved in designing, writing and scoring assessments. These all provide opportunities for professional development, especially if the work is on more complex performance tasks or portfolios.

States often involve some teachers in writing items, often multiple-choice, on state-made assessments, but scoring of writing samples and constructed response or performance tasks is often contracted out. It appears that few teachers are actually involved in writing a state's items, and often the writing is of multiple-choice items, which fails to provide substantial professional development for classroom assessment. A few states have made an effort to engage a wide range of teachers in writing performance tasks, and others have teachers involved in scoring.

Two cautions. First, good tasks and items are not easy to write, and learning to write them takes time. Therefore, rigorous quality review of items is necessary. Second, the time to do this work needs to be organized so as not to detract from teaching.

States often cite cost as the reason to contract out scoring. FairTest recommends that when costs are estimated, the value of professional development be factored in. It may well be that the narrowness of state writing samples, for example, renders them not good vehicles for professional development, whereas scoring portfolios and complex tasks has often been found to be a powerful form of teacher education. While we generally support having teachers involved in scoring at least the more extended constructed-response items, it may be that states find it more effective to use professional development funds in other ways.

Standard 4: Public education, reporting and parents' rights.

Parents and the public have the right to be informed about assessments and assessment results and to have access to all reports. Thus, reports at times will need to be prepared in languages other than English. When new assessments are introduced, extensive public education may be necessary. This is both fair to parents and likely to be vital to the success of new assessments. It is useful for states to find out what parents and the public most want to know and to make sure that reports are understood by their intended audiences.

Parents and students also should have the right to review assessments and challenge scores or items they believe to be flawed. *A cult of secrecy surrounds testing which serves to conceal its limitations from public understanding and mystifies students as to what high quality work looks like and what is wanted on tests.* Some states are making progress toward openness, but much more needs to be done. Openness is worth the cost of writing more items.

4.1. Parents and community members are educated about the kinds of assessments used and the meaning and interpretation of assessment results. Parents and the public deserve to know what kinds of assessments are used and why, and to have results of assessments reported in a clear and comprehensible manner. This includes how to interpret the results and important inferences that can be drawn from them.

States typically provide public reports, and many provide guidance on using the results, but few states appear to make an extensive education effort about assessment beyond publishing test scores. States introducing new assessments usually do try to inform the public about them. Some states release items or provide examples of items and student work. In reporting assessment results, states should also provide contextual information about the schooling students received, though as noted earlier no states said they did this. States also should clearly state the limits of the data and cautions about common misuses and misinterpretations.

4.2. The state surveys parents/public to determine information they want on assessments and whether assessment reports are understandable. Reports should include information that parents and the public want, and reports should be understood by audiences. This requires public opinion research.

Fourteen of the states responding to the FairTest survey reported surveying as to what information the public wants. Of those 14, six also surveyed as to whether the reports are understandable.

4.3. Reports should be available in languages other than English if a sizeable number or significant percentage of the student population come from homes where another language is commonly used. Spanish-language reports would be the most common.

Only five states reported that they reported in languages other than English. Many states with large numbers of LEP students did not provide such reports.

4.4. Parents and/or students have the right to examine assessments, appeal assessment scores, or challenge flawed items. Parental review encourages openness. States should release items or tasks on a regular basis. Because scoring can be incorrect and items may be flawed, clear processes for appeals and challenges are necessary.

Most states allow parents to examine tests, often under secure conditions, and a few release all or many items for public review after each administration. Review of commercial NRTs is more limited and difficult, but is allowed in some states, indicating that contractual problems with the testmaker (a reason some states cited for not allowing test review) can be resolved.

Eleven states reported on the FairTest survey that they allow item challenges or score appeals. Score appeals are more likely to be allowed on writing samples and constructed-response items, which are scored by people rather than machines, and on high school exit exams, where mistakes have more serious consequences.

Note: For a variety of reasons, some parents object to all or some kinds of large-scale testing. *Ten states reported allowing parents to exclude their children from an exam.* Some said requests for exemptions were growing, though the number remained small. A few even included the high school exit exam in the tests covered by such exemption policies, but in some of these the state said it would ask the parent to sign a form indicating awareness that the child would not receive a standard diploma if she or he did not take and pass the test. In such cases, given the relatively older age of the children and the consequences, it is probably wise for the child to also assent to opting out.

This was not an issue raised in the *Principles*. In the face of tests that may be more harmful than helpful, a parental right to exempt children may be reasonable. A caution should be raised, however, that schools do not use such a right as a lever to persuade parents of low-scoring children to opt out -- that is, to push them out.

Standard 5: System review and improvement.

States should regularly review their assessment programs in order to assure the quality of the system, to prevent or remedy harmful consequences of test use, to support beneficial consequences, and to provide information useful for improving the system. A comprehensive review would include the factors discussed in the *Principles*. This would include the quality and effectiveness of bias reduction, the extent of inclusion, professional competence in assessment, and the quality of public reporting. Including assessment as part of a review of a state's entire educational program probably makes more sense than just conducting separate reviews of assessment.

While most states conduct some form of review, their review practices are limited and important areas are often not addressed.

A comprehensive review of an assessment used for public information or accountability would help determine if:

- the data are accurate;
- the accountability system is relevant to important issues and actually reports what it says it reports (e.g., a report on writing is based on educationally valid understandings of writing);
- any impact the assessment has is at least neutral, preferably positive, and certainly not harmful to curriculum, instruction, student progress, or the cognitive and emotional development of children; and
- assessments measure in a balanced manner all important aspects of the standards or curriculum on which they are based and thus assess critical thinking and cognitively complex activity within and across subject areas.

Few states can provide data about their assessment program with respect to these key issues. In an era in which testing is proposed as a fundamental tool for school reform, states often can report little more than that scores are increasing or decreasing. They often cannot even be sure whether increasing scores are based on real learning gains or teaching to the test. Additionally, though most states have powerful leverage over district practices, such as through state constitutions, few states have evaluated their districts' assessment practices.

There is a further issue: the values and assumptions that underlie state reviews. For example, some states have concluded that the multiple-choice tests they use are appropriate for young children, contrary to the professional consensus in the field. Others claim that their multiple-choice tests can assess complex and critical thinking, which suggests that they and their critics may hold different conceptions of critical thinking.

We were able to examine a few independent and self-evaluations of states. The conceptual structures and values of the evaluators are clearly important in how they frame their approaches. Acceptance of traditional psychometric values and concepts, which underlie traditional exams, produce different evaluative conclusions than those based on different views of learning (such as constructivist or social constructivist models) or of the goals of schooling. *Reviewers need to make explicit and defend the perspectives, assumptions and values which undergird their reviews.*

Improving the evaluation process should be a priority in most states. The reviews must seriously and critically engage the underlying concepts of the state assessment programs.

5.1. The assessment system is regularly reviewed.

Twenty-eight of the forty-three states which responded to the FairTest survey reported that they have some sort of review process. All states should have comprehensive review procedures.

5.2. The review includes participation by various stakeholders and evaluation by independent experts. Participation by the public and independent experts helps ensure credibility and brings diverse views to the review process. While test developers or contractors should participate in evaluating the system, they are not independent evaluators.

Twenty-three states reported involvement by educators, 10 by one or more community sectors, 16 by SEA staff, three by test contractors. Three employed independent, outside experts. In general, the range of stakeholders involved is limited, and few states arrange for outside evaluation with any regularity, if at all. A few states have studied their systems in great detail and used outside experts as well as at least some stakeholders. These states are often those which have begun to develop fundamentally new assessment systems, such as Kentucky and Vermont.

5.3. The review studies how well the system actually is aligned to standards.

While some states reported studies as to the match between state standards or curriculum and the assessments, the reviews often fail to evaluate how well the assessment measures all aspects of the standards. In most cases, the studies appear to focus on whether test content is included in the standards; this is particularly the case when the match is to a commercially published test.

5.4. The review studies the impact of the assessment(s) on curriculum and instruction. Assessments can have a variety of consequences for school practice and the actual curriculum and instruction students receive. These consequences -- desired and undesired, beneficial and harmful -- should be studied in order to eliminate problems and enhance strengths.

Only 13 states reported studying the impact of state-mandated assessments on curriculum and instruction. Some states reported increased scores on the assessments as a positive impact. While teaching to the test can be positive if it does not narrow instruction in harmful ways, without further study states cannot be sure how much gain is real learning and how much is test-score inflation on a too-narrow test that is taught to in too-narrow ways.

5.5. The review studies whether assessments assess critical thinking or the ability to engage in cognitively complex work within a subject.

A mere five states reported studying whether the assessments measured critical thinking or cognitive complexity. Most state assessments are dominated by methods known to have limited capacity to assess critical thinking, but most states do not investigate this issue.

5.6. Reviews for assessments at grade 3 or below study whether the assessments are developmentally appropriate. Experts on the education of young children have advocated that assessment be "developmentally appropriate," that is, reasonable for the range of capabilities and ways of learning of students through age 8 (see Bredekamp).

Most states which test at or below grade 3 claim to have studied the assessments for developmental appropriateness, but it appears some of these studies may not include critical issues raised by experts on this age group. Of 24 states with mandated assessments at grade 3 or earlier, two reported studying them for developmental appropriateness (the actual number may be slightly higher, as not all states responded to the full FairTest survey). Guidelines for developmentally appropriate assessment for young children have cautioned against the use of multiple-choice tests, but some states have said they have reviewed their multiple-choice tests for appropriateness. It would appear, therefore, that those guidelines have not been used in selecting or evaluating the assessments.

5.7. Reviews study the impact of assessment programs on student progress and particularly the impact of any high-stakes tests, such as high school exit exams, on graduation rates. If graduation tests, for example, reduce the graduation rate or do so differently for different population groups, the state should know this and take appropriate steps to address the problem.

Seventeen states have mandatory high school exit exams. Of these, 12 responded to the FairTest survey and only four of them reported studying the impact on high school graduation. Since the use of single exams as a hurdle to high school graduation or grade promotion violates professional standards, states that persist in doing so should study the consequences of those exams. Preferably, the studies should be done by independent contractors not invested in the outcomes of such studies.

5.8. Reviews study the technical quality of assessments. Technical considerations, most importantly validity, but also generalizability, reliability, bias, and scoring procedures, should always be studied. Validity is fundamental, and overlaps with the topics addressed above, including the match with standards, assessment of critical thinking, impact on curriculum and instruction and on high school graduation rates, and bias. Gathering evidence about the validity of an assessment is a continuing process rather than a one-time effort.

Far too few states conduct technical studies of their assessments. Fourteen states reported doing technical studies. Technical studies on commercial tests are usually done by the publishers. Technical and consequential aspects of validity are complementary and both must be studied. This survey did not investigate the nature of the technical studies to determine what elements were included in the studies, nor was the quality of the studies evaluated.

5.9. The state reviews local assessment practices. This should include use of surveys regarding classroom, school or district assessment practices. This standard suggests that states have a responsibility to oversee district assessment practices in order to help prevent harmful practices and to support improvement.

Very few states survey to find out about district, school or teacher assessment practices, or review or evaluate local assessment practices. Four reported that they review district assessments, and one reported reviewing school assessments.

***5.10. Reviews help guide improvements in the assessment system that will bring the program more in line with the* Principles and Indicators.** Studies of the system should provide information useful for improving the system. The *Principles and Indicators* should be used to help shape the changes in a beneficial direction.

Few states that are revising their assessment systems reported using studies of the current or previous system in making revisions. Some state changes represent progress toward the Principles. *Others do not or are even steps backwards.*

B. Standards for Evaluating State Assessment Systems

Standard 1: Assessment supports important student learning.

1.1. Assessments are based on and aligned with standards.

1.2. Multiple-choice and very-short-answer (e.g., "gridded-in") items are a limited part of the assessments; and assessments employ multiple methods, including those that allow students to demonstrate understanding by applying knowledge and constructing responses.

1.3. Assessments designed to rank order, such as norm-referenced tests (NRT), are not used or are not a significant part of the assessment system.

1.4. The test burden is not too heavy in any one grade or across the system.

1.5. High stakes decisions, such as high school graduation for students or probation for schools, are not made on the basis of any single assessment.

1.6. Sampling is employed to gather program information.

1.7. The evaluation of work done over time, e.g., portfolios, is a major component of accountability and public reporting data.

1.8. Students are provided an opportunity to comment on or evaluate the instruction they receive and their own learning.

1.9. Appropriate contextual information is gathered and reported with assessment data.

Standard 2: Assessments are fair.

2.1. States have implemented comprehensive bias review procedures.

2.2. Assessment results should be reported both for all students together and with disaggregated data for sub-populations.

2.3. Adequate and appropriate accommodations and adaptations are provided for students with Individual Education Plans (IEP).

2.4. Adequate and appropriate accommodations and adaptations, including translations or developing assessments in languages other than English, are available for students with limited English proficiency (LEP).

2.5. Multiple methods of assessment are provided to students to meet needs based on different learning styles and cultural backgrounds.

2.6. Students are provided an adequate opportunity to learn about the assessment.

Standard 3: Professional development.

3.1. States have requirements for beginning teachers and administrators to be knowledgeable about assessment, including appropriate classroom practices.

3.2. States provide sufficient professional development in assessment, including in classroom assessment.

3.3. States survey educators about their professional development needs in assessment and evaluate their competence in assessment.

3.4. Teachers and other educators are involved in designing, writing and scoring assessments.

Standard 4: Public education, reporting, and parents' rights.

4.1. Parents and community members are educated about the kinds of assessments used and the meaning and interpretation of assessment results.

4.2. The state surveys parents/public to determine information they want on assessments and whether assessment reports are understandable.

4.3. Reports should be available in languages other than English if a sizeable number or significant percentage of the student population come from homes where another language is commonly used.

4.5. Parents and/or students have the right to examine assessments, appeal assessment scores, or challenge flawed items.

Standard 5: System review and improvement.

5.1. The assessment system is regularly reviewed.

5.2. The review includes participation by various stakeholders and evaluation by independent experts.

5.3. The review studies how well the system actually is aligned to standards.

5.4. The review studies the impact of the assessment(s) on curriculum and instruction.

5.5. The review studies whether assessments assess critical thinking or the ability to engage in cognitively complex work within a subject.

5.6. Reviews for assessments at grade 3 or below study whether the assessments are developmentally appropriate.

5.7. Reviews study the impact of assessment programs on student progress and particularly the impact of any high stakes tests, such as high school exit exams, on graduation rates.

5.8. Reviews study the technical quality of assessments.

5.9. The state reviews local assessment practices.

5.10. Reviews help guide improvements in the assessment system that will bring the program more in line with the *Principles and Indicators*.

C. Scoring Guide

The FairTest evaluation focuses on the primary characteristics described below. States' scores are based primarily on their current programs, but on occasion changes that are currently being implemented were considered.

Level 1. State assessment system needs a complete overhaul. Such a state system exhibits *three or more* of the following negative characteristics:

- Uses all or almost all multiple-choice testing;
- Tests all students in one or more grades with a norm-referenced test;
- Has a single exam as a high school exit or grade-promotion requirement; or
- Exhibits generally poor performance on the other standards.

Level 2. State assessment system needs many major improvements. Such a state system has *two* of the following negative characteristics:

- Uses all or almost all multiple-choice testing;
- Tests all students in one or more grades with a norm-referenced test;
- Has a single exam as a high school exit or grade-promotion requirement; or
- Exhibits generally poor performance on the other standards.

Level 3. State assessment system needs some significant improvements. Such a state system has some positive attributes but still has *one* of the following negative characteristics:

- Uses all or almost all multiple-choice testing;
- Tests all students in one or more grades with a norm-referenced test;
- Has a single exam as a high school exit or grade-promotion requirement; or
- Exhibits generally poor performance on the other standards.

Level 4. State assessment system needs modest improvement. Such a state system generally performs well across the standards, has none of the major problems described at previous levels, but does not show all the characteristics of a model system, including use of sampling and classroom-based assessments for accountability and public reporting.

Level 5. A model system. Such a state system performs well across all the standards, including use of sampling and classroom-based assessments as significant portions of accountability and public reporting. It may need minor improvements in some areas.

Not scorable. The state does not have an assessment system and does not mandate any assessments for districts to use, or is otherwise not scorable.

Discussion. This scoring guide gives the most weight to Standard 1. If an assessment system does not support high quality teaching and learning, it should be completely overhauled. The presence of some ameliorating characteristics such as limited use of NRT (e.g., only one grade and subject) or alternatives to the graduation requirement, or some other significant positive attributes from the other standards can move a state up a level.

D. STATE DATA TABLE

1996-97

STATE	level	m-c	nrt	grad test	writing	purposes
AL	1	1	1	1	1	1,4,6
AK	1	1	1	3	1	1,2,6
AZ	1	1	1			1,2,6
AR	2	1	1	3		1,2,6
CA*	2	2	1			1,5,6
CO	4	2			1,3	1,6
CT	4	2/3			1,3	1,2,6
DE**	0	4**		3,2	1	1
FL	1	1	1	1	1	1,2,4
GA	1	1	1	1	1	1,2,3,6
HI	1	1	1	1		1,2,6
ID	2	2	1		1	1,2
IL	3	1	1		1	1,4
IN	2/1	2	1		1	1,2,3,4,6
IA	0					
KS	3/4	2			1	1,2
KY	4/3	3	1		2	1,2,3,4
LA	1	2	1	1	1	1,2,5
ME	4	4			1	1,2,5
MD	3	3	2	1	1	1,2,3,4,6
MA	2	1	1	3		2
MI	3	2		4	4	1,2,3,4,5,6
MN	2	1		2	1	1
MS	1	1	1	1	1	1,2,4,6
MO^	4/3^	1			1,3	1,2,4,6

STATE	level	m-c	nrt	grad test	writing	purposes
MT	2	1	1			1,2
NE	2	1	1			2
NE	2	1	1	1	1	1,2,6
NH	4	2			1	1,2
NJ	2	2		1	1	1,2,4,5,6
NM	1	2	1	1	1,2	1,2,6
NY	2	2		1	1	1,2,3,4,5,6
NC	1	2	2	1	1	1,2,3,4,6
ND	2	1	1			1,2,4,6
OH	2	1		1	1	1,2,6
OK	2/1	1	1		1	1,2,4,5
OR	3	2			3	1,2,6
PA	3	2			1	1,2,3
RI	3	2/3	1		1	1,2,6
SC	1	1	1	1	1	1,2,3,4,5,6
SD	2	1	1			1,2
TN	1	1	1	1	1	1,2,3,4,6
TX	2	1		1	1	1,2,3,4,5,6
UT	1	1	1			1,2,5,6
VT	5	4			2	1,2
VA	1	1	1	1	1	1,2,5,6
WA	2	2	1		1	1,2,3
WV	1	1	1	4	1	1,2,4,5,6
WI	2	2	1		1	1,2,4,6
WY+	0	4				1

Coding and notes follow on next two pages.

Coding of table

level = the level of the state program according to the FairTest scoring guide
1 = needs a complete overhaul
2 = needs many major improvements
3 = needs some significant improvements
4 = needs modest improvement
5 = model system
0 = no state system and no state mandate for particular district testing; or otherwise not scorable

mc = multiple-choice, excluding writing assessment
1 = all/almost all m-c
2 = majority m-c
3 = minority m-c
4 = no/almost no m-c

nrt = use of a norm-referenced test (NRT)
1 = uses an NRT
2 = uses an NRT, but on a sampling basis

grad test = graduation test
1 = has a test and passing it is required for graduation
2 = has a required graduation test, but also an acceptable alternative
3 = state plans to require a graduation test but does not now have one
4 = has a graduation test, but passage is not required for diploma

writing = states have a writing assessment
1 = write to a prompt
2 = portfolio
3 = multiple choice
4 = anything else for writing

purposes = purposes for the test
1 = improve curriculum and instruction
2 = program evaluation/public reporting
3 = rewards for schools/districts
4 = sanctions for schools/districts
5 = rewards or sanctions for students other than high school graduation
6 = student diagnosis

Notes:

Data is from 1996-97 school year, except 1995-96 for Arkansas, Connecticut, Florida, Maryland, Mississippi, Ohio, which did not respond to FairTest survey.

In the "level" column, use of a slash (/), as in 4/3, indicates that the system is on the border; the first number is the direction in which the state appears to be leaning. In this column, numbers separated by a comma indicate a system whose parts (current, or current and being implemented) require separate evaluation.

In the multiple-choice ("m-c") column, use of a slash (/) indicates we could not precisely determine the proportions of multiple-choice items used on state assessments.

* California pays districts to test voluntarily, mostly with NRTs (hence a 2) and has other exams that are criterion-referenced with some constructed-response (hence a 3).

** Delaware assessed only writing 1996-97, not a full state testing program, hence a 0. Its new program is still being designed, but it will include norm-referenced tests and a high school exit exam (which will allow for alternatives) hence a 2.

^ Missouri's incoming program appears likely to score at a level 4; the current program, which relies primarily on criterion-referenced multiple-choice items but employs sampling, rates a 3.

+ Wyoming assessed only employment readiness in 1996-97, and that on a sampling basis, making it really a state without a state assessment system.

ALABAMA

Summary evaluation.

Alabama's testing program needs a complete overhaul, including: decreasing the use of NRTs and multiple-choice exams, increasing the use of multiple assessment methods, reducing the amount of testing, ending the high school exit requirement, strengthening bias review, making adequate provisions for LEP students, strengthening all aspects of professional development and making more efforts to educate the public. The state also should ensure regular review of the quality and impact of the state's assessments. FairTest hopes the review the state currently is undertaking will examine these areas and recommend major changes.

Standard 1: Assessment supports important student learning.

Alabama has content standards in math, science, English language arts, social studies, foreign languages, music, visual arts, health education, physical education, vocational education and drivers' ed. Performing arts is in development. Each is revised about every six years. Committees mainly of educators develop the standards, which are subject to public comment prior to Board of Education adoption. The standards are used to guide test development and textbook adoption.

Alabama's state assessment program administers the Stanford 9 NRT annually to students in grades 3-11 in the areas of language, math, reading, science and social studies. The criterion-referenced, multiple-choice Basic Competency Test assesses students in grade 9 in the areas of language, math and reading. The High School Basic Skills Exit Examination is a multiple-choice criterion-referenced test in language, reading and math given beginning in grade 11. Math end-of-course tests in Geometry and Algebra are administered when a student takes the course. These are criterion-referenced and include multiple-choice and constructed-response items. For these components, all students in designated grades or subjects are tested and all students see the same items.

The state has a writing assessment in grades 5 and 7 that uses writing samples with SEA provided prompts. All students are assessed. Fifth graders respond to one of three prompts (they do not choose which one) and receive 50 minutes to produce a sample on demand. Seventh graders select from one of four prompts and have 60 minutes. Scoring is done by a commercial company using state-developed scoring guides.

The Differential Aptitude Test with Career Interest Inventory (an off-the-shelf, norm-referenced, multiple-choice test) has been offered to students in grade 8, but is being discontinued this year. Alabama recently made mandatory the previously optional kindergarten testing, and the SEA is now deciding what test(s) to use. The intended purposes are to provide information to teachers for use in instruction and early academic intervention, and to inform pre-kindergarten program improvement.

The SEA reports that some areas in the state standards, such as speaking, are not assessed. Time, money and use of multiple-choice items also limit what can be assessed. When tests are developed in Alabama, content standards are used as the basis for item development. Course content is set by a state committee and includes higher order thinking in the subjects. Tests are reviewed by teachers for content match. The Algebra and Geometry tests were developed in part to encourage higher order thinking in these subjects in the classroom. For the exit exam, an instructional validity study is done, including surveys of teachers and students.

The Basic Competency Test and the High School Exit Exam are used for student diagnosis or placement or both. All components (except DAT) are intended to be used to improve curriculum. All assessments are used in school performance reporting. High-stakes accountability consequences for schools and districts (warnings, probation/watch lists, and takeover) are based in part on the NRT. The results of the math end-of-course test can be included in determining the course grade at LEA discretion.

Evaluation: Alabama has standards and uses them to develop the state assessments. The involvement of educators in developing and reviewing assessments is positive, as is the survey undertaken for instructional validity on the high school exit test. Alabama is weak on multiple methods, relying almost entirely on multiple-choice, particularly the NRT. It uses a single test as a graduation requirement and it imposes too many tests, particularly the use of the NRT in grades 3-11. Overall, Alabama's testing program contains too many negatives relative to the positives to be viewed as supportive of important learning. It should move away from the NRT, possibly by reducing it to a light sampling tool as does North Carolina, and make its CRTs mostly constructed-response. It should also cease use of a high-stakes graduation test.

Standard 2: Assessments are fair.

For state-made tests, a bias review committee looks at a range of bias and equity issues, including special educational needs, and makes recommendations for improvement. The SEA uses this information in developing the pilot assessment. Statistical reviews are performed on the pilot assessments and the data used for further refinements.

About 8 percent of elementary students tested are classified LEP or IEP. All decisions concerning inclusion of special needs students are made by IEP/LEP/504 committees. (504 includes students with disabilities who are not in special education and do not have a special curriculum.) For IEP, 504 and LEP, some students are exempted from taking tests, including from the HS exit test, but to earn a diploma these students do have to pass the HS exit test. A variety of accommodations are used with IEP and 504 students (which vary by test), while there are no accommodations for LEP students. This year, IEP students are not included in regular reports for some tests, are included in others (NRT, end-of-course math, writing), while LEP and 504 who are tested are included. The plan is to include all tested students in regular reports next year.

Evaluation. The bias review effort could be strengthened by providing the review committee with more authority. IEP appears acceptable except for the graduation requirement. LEP is not adequate. Moving to include all tested students in regular reports is positive. Heavy reliance on multiple-choice hinders equity by not reflecting the needs of diverse learning styles or cultural backgrounds. Demographic data are not reported. Overall, the state is moderately weak on this standard.

Standard 3: Professional development.

The SEA provides print materials to educators and policymakers, and video materials to teachers, as part of professional development. Packets of annotated papers from the writing assessment are distributed to teachers for instructional use. Alabama has no particular requirements for pre-service teachers, while in-service teacher requirements for professional development focus on using standardized test results. The SEA has done only a little on professional development for classroom assessment, mostly at LEA request. The state has not evaluated teacher assessment competence or surveyed educators to determine their professional development needs.

Evaluation. The state appears to have good involvement of educators in writing items and evaluating the state-made assessments. Otherwise, the state is weak on professional development in assessment for teachers, save only the writing assessment packets. Teachers do not score the writing samples.

Standard 4: Public education, reporting and parents' rights.

Alabama has a State Superintendent's Report Card reporting data at the district level, and beginning next fall also at the school level, to the media, legislators, districts and schools, in English only. Students or parents can appeal a score or challenge items on state-made exams, but there is no mechanism for parental review of items or tests.

Evaluation. Public information is only basic. Parents and the public have not been surveyed to determine their needs and understanding. Parents' and students' rights are too limited.

Standard 5: System review and improvement.

A test advisory committee is now reviewing the state assessment program. It will look at the impact of the tests on curriculum and instruction. The state has not surveyed or evaluated district assessment practices.

Evaluation. It is positive that the state is now conducting a review, but more should be done regarding evaluation of district practices, the assessment of critical thinking, the impact on high school graduation, and developmental appropriateness. Whatever assessment is chosen for the now-mandatory kindergarten test, it should be developmentally appropriate and should be carefully monitored to determine its real impact on local practices and consequences for children. Continuing evaluation needs to be built into the system.

Alabama responded to the short form of the FairTest survey via a telephone interview. We also used CCSSO/NCREL, CCSSO and AFT reports. Alabama reviewed a draft of this report.

ALASKA

Summary evaluation.

Alaska's assessment program needs a complete overhaul. Planned changes do not solve existing problems, particularly the reliance on an NRT -- the CAT. In addition, they add a major problem -- a mandated high school graduation test. The state plans to develop a system based on standards, but does not yet have such a system, relying instead on an NRT plus a writing sample, piloted this year. The revisions propose retaining the NRT and adding state assessments in other grades. This will produce too high a test burden. The NRT should be dropped, or at most retained on a sampling basis, as it will be inadequate for assessment based on high standards for a multi-cultural population and because it utilizes only one method (multiple-choice). New assessments should be primarily performance assessments. The high school exit requirement should be dropped before it is implemented. In terms of equity, the major concern is with fair assessment of the state's large native population. Professional development should be substantially expanded to support improved classroom assessment and new state assessments. The evaluation system also needs strengthening.

Standard 1: Assessment supports important student learning.

Alaska has recently approved formally voluntary state standards in English language arts, math, science, government/citizenship, history, geography, healthy life skills, world languages, technology and the arts. Curriculum frameworks and professional development are being developed in most of these, and future state assessments will be based on them. Standards were developed by committees of educators and public representatives and were subject to a public review process.

The SEA currently administers off-the-shelf, norm-referenced tests (CAT/5) in grades 4, 8 and 11. A writing assessment component using samples in response to SEA provided prompts was piloted in 1996-97 for grades 5, 7 and 10. It is voluntary for students, but will be mandatory for districts to administer next year. The pilot has been aligned with state language arts standards. For the NRT, the publisher reportedly aligned test items with state standards. The pilot writing assessment was scored by trained state teachers using a rubric designed by teachers, administrators, SEA staff and outside experts.

The state is developing a new assessment plan which will see students in grades 3-11 tested by one or another test (CAT, writing, other state exam) each year, though no funds have yet been allocated. Math exams for various grades using multiple-choice and short-answer constructed-response items are being piloted. The new assessments will be aligned to the standards.

This spring, the state mandated a high school exit exam. No specific consequences for schools are attached to assessment results, but some may be in the future (e.g., made part of accreditation). Results of statewide assessments also are used for student diagnosis or placement, curriculum/instruction improvement and program evaluation. A student questionnaire is attached to the NRT.

Evaluation. The current assessment program is light but, being essentially an NRT plus a writing sample, inadequate and misdirected. Rather than develop an overly burdensome plan utilizing both an NRT and state assessments, Alaska should replace the CAT with a state assessment based on the standards. This assessment should use a minimum of multiple-choice items and mostly use a mix of short- and extended-response and performance items. At most, the NRT should be used on a sampling basis for program information. The adoption of the high-stakes graduation test should be dropped. The questionnaire now attached to the NRT should be retained.

Standard 2: Assessments are fair.

Bias committee members, selected for geographic, ethnic and gender diversity, can recommend changes in the state-made assessments. Writing sample prompts were reviewed for bias prior to assessment, and plans are to review them again after administration. Prompts and the writing process were intended to accommodate students from different cultures and with different learning styles. Bias review is done by the publisher for the CAT/5.

Currently, 8 percent of students who are tested have an IEP and 8 percent of tested students are LEP. An unknown percentage of IEP are not tested. The GWU Center for Equity reported that about 21 percent of the state's students are LEP, most speaking indigenous languages (there are about 85 indigenous languages spoken in Alaska, some by a few hundred people, and some are not written). If these numbers are accurate, either many students are not tested, or they may be mis-tested in English. However, by grade 4 most of Alaska's LEP students have been in school learning English since kindergarten or grade 1 (though for some groups, school participation can be erratic). The state is revising its methods for collecting data on who is assessed and who is exempted. Results for those who are tested are included in overall state data. Thirteen allowable accommodations for the CAT/5 were pilot-tested this spring; the state expects most to be approved.

Evaluation. Bias review appears to be adequate, though the authority of the review committee may need to be strengthened. One major problem appears to be how the state can assess adequately its large percentage of LEP students, or at least be sure that the students are sufficiently proficient in English to be assessed in English. Use of the CAT/5 with its limited format is an obstacle for assessing a culturally varied population and for assessing students with different learning styles. New assessments should allow multiple means to demonstrate achievement. Data should be reported by demographic category. Mandating an exit exam also runs contrary to this standard.

Standard 3: Professional development.

Various professional development opportunities are offered by the state, but it has no requirements for pre-service or in-service teachers. It does not evaluate teacher competence in assessment or survey educators for their professional development needs. Examples and scoring guides for the writing assessment are available to teachers and administrators. Summer institutes on the writing assessment were held for teachers. Educators participated in developing and scoring the writing assessments.

Evaluation. A substantial increase in and systematization of professional development is needed, focusing on classroom assessment capability and including pre- and in-service teachers. Educator involvement in developing and scoring assessments should be maintained and expanded.

Standard 4: Public education, reporting and parents' rights.

Examples and scoring guides for the writing assessment are available to students. Parents or students can request to see the test after administration. For the NRT, only grade 4 practice materials exist. Some materials on the test are sent to parents. Assessment results are reported in two to three months, in English only, with guidance on use of the results. The public has not been surveyed as to information it wants or whether the reports are understood.

Evaluation. As new assessments are developed, strong communications and a policy of openness toward the public and toward providing examples to students, parents and the public will be important. Reporting in languages other than English should be considered.

Standard 5: System review and improvement.

The State Board of Education, including administrators and representatives of the general community, annually evaluates the state assessment program. District assessments are not evaluated, but the state has surveyed district practices. Though state assessments are intended to guide curriculum and instruction, the impact of assessment on them is not studied (note that the writing assessment is brand new).

Evaluation. A stronger evaluation process will be needed as a new assessment program is introduced, including studying the actual impact on curriculum and instruction, whether the assessments are fair to all students, and whether they fully assess the state standards. Information should be obtained and used to steadily improve the system.

Alaska responded to the full FairTest survey and answered additional questions by telephone interview. We also used CCSSO/NCREL, CCSSO and AFT reports. Alaska reviewed a descriptive draft of this report.

ARIZONA

Summary evaluation.

Arizona's system needs a complete overhaul. It marks a major step back from the previous system, which relied heavily on performance assessment. The state should drop its NRT requirement in grades 3-12 and re-establish a program that includes support for districts to use performance assessments. The state is weak on fairness and professional development.

Standard 1: Assessment supports important student learning.

Arizona has had, since the early 1990s, Essential Skills curriculum guides in language arts, math, science, social studies, health, foreign language, literature and several performing arts. The state is now developing content standards in nine areas: language arts, math, science, social studies, arts, comprehensive health, foreign language, technology and workplace skills. New assessments and a review of local assessment requirements to align them with the standards will be undertaken when standards are completed.

In the spring of 1997, the newly adopted Stanford 9 NRT was administered by the state to all students in grades 3 through 12. In contrast, in the fall of 1995, Arizona administered only the multiple-choice NRT portion of the state assessment program (customized ITBS) to grades 4, 7 and 10 in language arts, math and reading. The state-level performance-based program (Arizona Student Achievement Program -- ASAP, used from 1992-94) is currently on hold, pending decisions about new assessments. It is likely that it will not be re-installed. The ASAP appears to have been the victim of competing political and educational agendas, both during its development and in the decision to suspend it (see Smith in bibliography).

Results of the Stanford 9 are used for student diagnosis or placement, curriculum improvement and program evaluation. They are also used for school accountability via a school report card. Any other uses are at district discretion.

In addition to the state tests, Arizona requires districts to assess their students, but the requirements allow substantial flexibility as to how districts do this, and as a result, district assessments vary greatly. ASAP had a requirement that districts develop a set of constructed-response or performance-based assessments in reading, writing and math to be used in determining high school graduation. They would be scored locally using the state's rubric. As the state's program is being revised, what LEAs will be required to do remains undetermined.

Evaluation. Use of an NRT in grades 3-12 represents a significant step backward for Arizona. The state program is now too heavily multiple-choice, is norm-referenced, and tests far too many grades. Positively, the stakes are relatively low. ASAP was a significant experiment in the use of state performance assessments. It was intended to support improved classroom instruction and assessment and to provide accountability. Independent evaluations suggested that the purposes were contradictory and the accountability uses not well worked out (see Smith). Still, by being performance-based and in only a few grades, it was a substantially superior approach. The state needs to repeal use of the NRT and re-establish a performance assessment program. A focus on supporting strong LEA assessments should also be re-established.

Standard 2: Assessments are fair.

Students with limited English proficiency (LEP) are exempt for 3 years from the NRT. IEP students may be exempted based on their plan (no numbers were provided). LEP and IEP students who are tested are included in regular reporting.

Evaluation. An absence of data makes this hard to evaluate. Most NRTs allow too-limited accommodations for either IEP or LEP and lead to weak inclusion. The emphasis on multiple-choice does not meet the *Principles and Indicators* requirement for variety in assessment methods.

Standard 3: Professional development.

The state offers no professional development in assessment, though professional development was attached to the performance assessment system.

Evaluation. The independent reviews concluded that professional development attached to ASAP was quite inadequate, but it did exist. The state needs to support teacher competence in classroom assessment and provide pre- and in-service professional development.

Standard 4: Public education, reporting and parents' rights.

Standard 5: System review and improvement.

No data were provided to respond to these two standards.

Arizona wrote a one-page letter rather than respond to the FT survey. CCSSO/NCREL, CCSSO and AFT reports were used. The state reviewed a descriptive draft.

ARKANSAS

Summary evaluation.

Though Arkansas did not respond to the FairTest survey, it appears from other data that its assessment system needs many major improvements. It uses an NRT and is implementing a CRT, both all multiple-choice except perhaps for a writing sample on the CRT. The testing burden is becoming fairly heavy with a full-battery NRT and the CRT, though the tests are administered in different grades. The state also is planning a high school exit exam. The state should drop the NRT or use it only on a sampling basis, not implement the exit exam, and change the CRT to become primarily a performance assessment. Bias review data were not reported, and inclusion with proper accommodations of IEP and LEP students needs improvement, particularly for LEP. Positively, professional development appears to be fairly extensive and focused on performance assessment. No data were reported on parental rights or on state system review.

Standard 1: Assessment supports important student learning.

Arkansas has content and performance standards in math, reading and English language arts, which have been implemented. Standards for science, foreign languages, social studies, fine arts, health and physical education are being developed.

The state uses a commercial NRT, the Stanford 8, in grades 5, 7 and 10, in language arts, math, reading, science and social studies.

The state is implementing, with a contractor, a CRT to be used at grades 4, 8 and 11/12, in math, reading and writing. Pilot testing has been done for the grade 4 and 11/12 tests. No detailed information was available on the writing exam, but it probably involves responding to a prompt. The grade 11/12 exam will be a high school graduation test, but no date has been set for when passing this test will become mandatory for graduation.

Districts are required to assess in grades K-4 using multiple measures and students not functioning on grade level in reading or math must attend summer school or be retained.

The state is not developing any non-multiple-choice items.

The tests are used for student diagnosis, improvement of curriculum and instruction, and program evaluation. The Stanford is also used for school performance reporting.

Evaluation. Arkansas' assessment program needs major changes to meet this standard, including ending heavy reliance on multiple-choice and norm-referenced testing. The state should begin to use performance assessments. The planned graduation test requirement should not be implemented. No details were provided on the requirement for districts to assess in grades K-4, but the high stakes, including possible grade retention, require that the state pay careful attention to the assessments and how they are used.

Standard 2: Assessments are fair.

Print materials have been provided to students to prepare them for the tests. Data are reported at the state level on the CRT by race, gender, free lunch eligibility, IEP and LEP status, but only by race and gender on the NRT.

Students may be exempted from testing based on their IEPs. LEP students are tested unless a note from a parent in the native language requests the student not be tested. Fairly extensive accommodations are available for IEP students on the NRT, but not many on the CRT. Few accommodations are available for LEP students on either exam. Scores of some IEP and LEP students were not included in regular reports.

Evaluation. No information on bias review was provided. Inclusion with proper accommodations needs improvement for IEP students and even more for LEP students.

Standard 3: Professional development.

Print, video, television and computer materials have been provided to educators for professional development. The SEA conducted statewide training of trainers in performance assessment, including performance tasks, projects, portfolios and direct writing prompts. With the contractor, they also conducted workshops on the state tests.

Evaluation. The trainer or trainers approach in performance assessment is a very positive sign, though the data provided do not enable us to evaluate the extent to which this program has reached teachers or whether it has had any effect.

Standard 4: Public education, reporting and parents' rights.

Print materials have been provided to parents about the test, and policymakers have also received video materials.

Evaluation. No data were provided on parental rights.

Standard 5: System review and improvement.

No data were reported.

Evaluation. In addition to monitoring the state assessments, the heavy requirements for districts to test in the early grades, including high stakes, should be carefully studied.

Arkansas did not participate in the survey. This report used two years of CCSSO/NCREL reports, plus the AFT report.

CALIFORNIA

Summary evaluation.

California's assessment program needs at least some significant and perhaps many major improvements. The state has taken major steps backwards in the past several years, and battles between the governor and the legislature are continuing about future testing.

The Pupil Incentive Testing Program (PITP), which pays districts to use an off-the-shelf test in grades 2-10, is becoming the de facto state testing program. At least 60 percent of the districts have agreed to participate, including over half the state's students. However, this program relies almost entirely on multiple-choice NRTs that are not tied to state standards. A few approved tests use a mix of methods. The program also tests in too many grades. The only positive here is that the stakes are not high, though results are reported.

Other state exams use a mix of methods and are more reasonable. A new state assessment system that will use a mix of methods and employ sampling in a few grades has been authorized but not developed. It is possible that PITP will be replaced by a requirement, being pushed by the governor, to test all students with one NRT; his opponents argue that such a test will not be aligned with state standards and the state should defer action until the new state assessment system is developed.

The state should at a minimum drop the requirement that for a district to receive any reimbursement in the PITP it must test all students in grades 2-10. It should not adopt the governor's proposal. It should develop the new system as planned, keeping stakes low, and it should re-establish its program for supporting local assessment development with substantial professional development support, an approach which runs counter to PITP. All other areas need strengthening, though teacher involvement in the other state exams seems reasonable.

Standard 1: Assessment supports important student learning.

The state has curriculum frameworks and is now developing content and performance standards. Frameworks were developed by teachers, curriculum experts and community members. Standards are being developed by an appointed commission with input from parents and other community members.

The PITP started in the spring of 1996. To become eligible for a payment of $5 per student tested, districts must select achievement tests covering reading, spelling, written expression and mathematics from a list of state-approved published exams. The tests are to be reviewed for various criteria, but in 1996 the key criterion was reliability. Districts will report results to students, teachers, and parents, and they will provide a summary of results annually to their governing boards. The tests are intended to guide curriculum and instruction.

Currently, the California statewide assessment program is limited to the Golden State Examinations (GSE) and the Career-Technical Assessment Program (C-TAP). The state plans to develop a new state assessment for grades 4, 8 and 10, using a mix of methods, to produce school-level data.

The purpose of the GSE program is for awarding honors diplomas and for improvement of curriculum and instruction. The GSEs are criterion-referenced, end-of-course exams offered on a voluntary basis to students in grades 9-12 in a variety of subjects. Districts must ensure students have an opportunity to take these tests. Curriculum frameworks are used as the guidelines for test development, and the SEA says the tests cover all general areas within the tested subject. They employ various combinations of enhanced multiple-choice, short or extended constructed-response, and individual performance assessment. The GSE was developed by teachers, administrators, SEA staff, outside experts and education organizations.

The C-TAP is used primarily to determine student readiness for employment by assessing workplace skills. It is now being extensively revised. It has used a combination of writing samples, performance testing (projects and student presentations) and portfolio assessment, scored at the local level using guides developed by the SEA in collaboration with teachers. While this continues, a contractor, WestEd, is piloting and field-testing new exams that will combine multiple-choice and short-answer response methods. These will be similar to the end-of-course GSE exams and will be scored at the state level. The new exams will be optional for high school students who want extra recognition. Both the performance assessments and the on-demand exams may be used in the future for certification of students in school-to-work programs.

Evaluation. Until vetoed by the governor in 1994, the state's assessment program was the innovative California Learning Assessment System. It used a mix of constructed-response and multiple-choice items and was perhaps the most controversial state exam in the nation. It represented a significant step forward in state testing as it used multiple methods and was also available in Spanish. CLAS included not only the exam but also development of local performance and portfolio assessments. The governor placated conservatives who objected to using anything other than multiple-choice items and vetoed the reauthorization. The legislature then adopted the PITP plus plans for a new assessment that would produce school-level data. While PITP is not officially a state assessment and is not mandated, it is becoming a virtual state assessment. Unfortunately, it is a wholly regressive program, relying mostly on multiple-choice NRTs and testing far too many grades with assessments not based on the state's standards. It remains to be seen whether the new assessment is created, and if so, what form it takes. The worst scenario would be replacing the PITP with a single mandatory NRT. Dropping the performance elements of C-TAP is likely to be a step backward for that exam.

Standard 2: Assessments are fair.

Bias and community review processes exist for both GSE and C-TAP exams. Committee members are chosen to reflect the statewide population in proportion to each demographic category. The committees include all stakeholder groups except students. Industry representatives review items for C-TAP. Committees have power to reject items or suggest revisions. Items are field tested and reviewed statistically for bias. For the PITP, any bias review is done by the testmakers. A state committee reviews tests submitted for PITP approval for items that ask for personal information, which they require publishers to delete.

About 11 percent of the state's public school students have an IEP, and 23 percent are LEP. IEP and LEP students are generally not exempted, as per state mandate.

Evaluation. The bias review procedures for GSE and C-TAP are fine. While CLAS had a Spanish version, it appears unlikely that the tests used in the PITP are appropriate for many

IEP or LEP students. Accommodations are left to the districts and the test publishers. This, plus the reliance on multiple-choice in PITP, makes the state weak on addressing bias.

Standard 3: Professional development.

Teachers are required to learn observational techniques and psychometrics in pre-service. The GSE provides information to teachers, students, parents and the community on format, content, sample questions, scoring guides and examples of student work and performance levels. No other professional development is provided. C-TAP plans to provide samples, etc., but the new exams are still in development.

Evaluation. Professional development needs strengthening, though the pre-service requirement for observational techniques is positive. A fairly extensive professional development program had been related to CLAS. By using commercial exams, PITP discourages the development of local assessments, and most of the PITP tests are not compatible with the kinds of approaches, mostly portfolios and performance assessments, associated with the local assessments developed under CLAS. This has in some instances created serious contradictions for teachers, which would make professional development more complicated. Teacher involvement in the GSE appears positive.

Standard 4: Public education, reporting and parents' rights.

Parents may request in writing that their children not be tested. Results of the GSE are reported 4 to 6 months after being administered to students, parents, schools and the public. Districts are required to make PITP test results public. It is not known how many do so in languages other than English.

Evaluation. The demise of CLAS stemmed in large part from the mobilization of some parents against the program. Studies of CLAS have suggested a lack of public education about the exam was a major failure. Any new state assessment that contains constructed-response items may face a similar challenge, and public education will be necessary. Under CLAS, some districts involved parents and community members with their assessment programs, but these are not statewide. As a large percentage of the state population does not have English as a first language, reporting and education should be done in other languages.

Standard 5: System review and improvement.

The state has surveyed district practices. GSE is intended to guide curriculum and instruction, but no evidence is gathered on its effects. The PITP is not aligned with state standards and most of the approved tests will not measure large areas of the standards.

Evaluation. Strengthened review is necessary, both of the state exams and of what happens at the district level with the PITP.

California responded to the full FairTest survey. We also used CCSSO/NCREL, CCSSO and AFT reports. California reviewed a descriptive draft of this report.

COLORADO

Summary evaluation.

Colorado's just-introduced and still limited state assessment program appears to be headed in a direction that will leave it needing modest improvement. On the two existing tests, reading and writing in grade 4, approximately one-half of the scoring is based on multiple-choice items, substantially more than it should be but less than in most states. Writing only to prompts and using multiple-choice items in writing presents a narrow picture of writing. Since the state is beginning a new program, it should do more with portfolios and extended-response tasks. Positively, stakes are relatively low and norm-referencing is not used. The plan to help districts with assessment development could be very positive since most districts in the US rely heavily on commercial NRTs. Bias review, inclusion and reporting are solid. Professional development may need to be expanded, particularly with a focus on classroom assessments, not just the state exams. Public education may not be sufficient for new assessments. The new program is too early to review, but the alignment study and teacher survey that have been done are promising signs of a future comprehensive review process.

Standard 1: Assessment supports important student learning.

State Model Content Standards were adopted in 1995 in reading/writing, math, science, history and geography, and performance standards have been developed in these subjects. Additional standards are being developed in art, music, civics, economics, foreign languages, and physical education. Districts were required to adopt their own content standards by January 1, 1997. Districts are responsible for designing their own curricula to reflect their own standards, which must meet or exceed the state standards.

The Colorado Student Assessment Program (CSAP) is criterion-referenced and aligned with state standards. It is just starting and will use multiple-choice, short and extended constructed-response items and performance tasks.

The spring 1997 administration was the first. All grade 4 students are assessed in reading and writing. On both of those tests, about half the score is from multiple-choice items. Students also respond to a state-determined prompt. Constructed-response items and writing samples are scored by the contractor, not teachers, using rubrics the contractor developed.

All items on the assessment were selected or developed to measure student performance on the state standards. Alignment to the content standards was evaluated by the SEA, the private test contractor, and committees of Colorado educators to ensure that all aspects of the reading/writing standards, except speaking and listening, are tested.

Additional grades and content areas will be phased in according to pending legislation. A grade 3 reading test will be developed. Rules have been established on developmental appropriateness for standards and criteria for the assessment of literacy for all students in grades K-3. The state also plans to help districts develop assessments based on local standards.

The assessment is intended to be used by schools to improve instruction and promote student progress in meeting the state standards. At this time, there are no consequences for students or schools attached to the exams. Surveys were sent to teachers, principals and district assessment coordinators after the assessment.

Evaluation. The exams are still too dependent on multiple-choice and limited writing samples. Portfolios should be considered for writing and perhaps other subjects. Teachers should be involved in scoring the extended-response and performance items and writing. The state reports it will use performance tasks as well as extended-response, but it appears that performance-tasks are not yet in use. The grade 3 reading exam will have to be carefully studied for developmental appropriateness and impact on curriculum, instruction and tracking practices.

Standard 2: Assessments are fair.

The SEA reported trying to take into account the variety of cultural backgrounds and learning styles of the student population. Items and tasks are pre-tested and analyzed for bias and will be analyzed after administration. A bias review committee representing the state population had the authority to delete or modify items in the item pool prior to the construction of the test forms.

Decisions as to whether to test IEP or LEP students are made on an individual student basis. The state developed guidelines for decisions about which students should be tested and what accommodations, if any, should be used. For the first year only, a commercial Spanish-language test, CTB's *SUPERA*, was used. In the future, assessments will be translated into Spanish, except possibly the reading exam. Reports are disaggregated by race, gender, LEP and IEP at the state and school levels, and by SES at the state level only. All those who take the test are included in regular reports.

Evaluation. Bias review seems generally strong, though the committee should have heavier representation by minority populations to include an adequate variety of perspectives. Use of Spanish-language tests will be a strong step forward, though translations, rather than developing an alternative assessment, can be difficult and flawed. Reporting disaggregated data is positive. Use of multiple methods should help better assess across variability in student learning styles and cultures.

Standard 3: Professional development.

Colorado has no pre-service requirements for teacher preparation in assessment. The SEA conducted assessment training sessions for district assessment coordinators, who were in turn responsible for training teachers. The state does not evaluate teacher competence in assessment. The SEA does survey educators to determine if their professional development needs are being met, and it plans to survey educators to determine the effectiveness of the new assessment materials.

For the new assessments, a Demonstration Book was prepared which contained examples of items and presented assessment guidelines. In the future, 25 percent of items will be released annually for professional development.

Evaluation. Professional development needs to be strengthened for pre-service and in-service teachers. If the state is to help districts develop new assessments, it should also help ensure that teachers can use them appropriately and have strong classroom assessment skills. The surveys are a positive step and should be continued, for example to see if district assessment coordinators do provide sufficient training. Evaluation of district assessment programs and teacher competence in assessment should be undertaken as part of review processes.

Standard 4: Public education, reporting and parents' rights.

Parents and the public have not been surveyed as to what information they want in assessment reports. Parents or guardians can refuse to have their children tested. As there are no direct consequences to students and the assessment is new, no policy has been developed regarding appeal of scores; however, individual items can be challenged. Parents or students cannot review items after completion of the assessment, based on the contract with the test developer. Reporting of spring tests will be in the fall, only in English. Public education about the new assessments was done by using the media and through availability of the Demonstration Book.

Evaluation. Public education perhaps should be strengthened; we do not know how accessible and widely distributed the Demonstration Book actually is. Parents should be able to review the test at their child's school under secure conditions, a policy allowed in some other states that use contractor-made or commercial tests. Reporting should be in Spanish as well as English.

Standard 5: System review and improvement.

Since this is the first year of the new assessment, the state has not yet evaluated assessment practices at the district, school or classroom level. A continuing review will be conducted as the system is developed, but details of what will be evaluated and how have not been finalized. Alignment was reviewed for the grade 4 reading and writing test, as was the ability of the test to elicit and assess cognitive complexity and critical thinking.

Evaluation. As a new system, initial review procedures have been reasonable. More comprehensive reviews, including of district practices, are needed in the future.

Colorado responded to the short form of the FairTest survey. CCSSO and AFT reports were used. The state reviewed a draft description.

CONNECTICUT

Summary evaluation.

Though Connecticut did not respond to the FairTest survey, it appears from other data sources that the state assessment system needs only modest improvements. It does not use norm-referenced testing and has no high stakes. The state relies heavily on constructed-response and performance items, though more so for the high school than elementary school exams. The performance parts of the elementary school assessments and perhaps also the high school exam should be expanded. Portfolios should be considered, as should sampling. It appears that inclusion of IEP and LEP students needs improvement. No data were provided on bias review. The information on professional development was limited, but some seems to be available, perhaps a good deal. Some form of public reporting is done, but no information was available on public education. There was no data on system review.

Standard 1: Assessment supports important student learning.

Connecticut's "Common Core of Learning" articulates the state's goals. Though content standards and curriculum frameworks were reported to be under development, performance standards and assessments were reported as completed. These are developed by the SEA with committees representing business, higher education, political and general education concerns. The SEA says that the assessment program is aligned to state goals and curriculum frameworks, but the AFT reports that curriculum guides are being revised and that the state assessments are not directly related to the existing guides.

Connecticut's assessment program includes the criterion-referenced Connecticut Mastery Test (CMT) given every fall to grades 4, 6 and 8 in the areas of language arts (reading, writing) and math. It contains mostly multiple-choice items. The reading test includes a customized Degrees of Reading Power test. Math has some short-response items, reading contains short- and extended-response and performance tasks, and writing includes responses to SEA prompts as well as multiple-choice items. The Connecticut Academic Performance Test (CAPT), also criterion-referenced, is administered in grade 10, in language arts, math, science and interdisciplinary topics. It is primarily constructed-response, with some performance assessment (reading and science) and some multiple-choice.

In writing, students in grades 4, 6, and 8 are allowed 45 minutes to produce a sample, while grade 10 students receive 90 minutes as the test involves a great deal of reading. Scoring is holistic using a state rubric designed by a commercial company. Writing is also assessed in the interdisciplinary assessment, and the SEA has developed research paper exercises for writing assessment.

Results of tests are used for student diagnosis or placement, improvement of curriculum, program evaluation, and staff accountability. The CAPT is used to award a high school skills guarantee/certificate of mastery to students. CMT results also have the possible consequence of funding increases for schools.

Evaluation. Connecticut performs fairly well on this standard as it includes a variety of constructed-response and performance tasks, does not use norm-referencing, and does not have high stakes. We do not have information on the proportions of time or scores allotted to the different methods. An increase in the proportion of constructed-response and performance tasks may be warranted, particularly on the CMT. Sampling and portfolios should become part of this program.

Standard 2: Assessments are fair.

Print materials are available to students to explain the assessments.

Nearly 14 percent of students tested have IEPs and about 1 percent of students tested are LEP. IEP and LEP students may be excluded. A variety of accommodations can be allowed for IEP, none for LEP. Scores of all tested students are included in regular reports and evaluated according to regular state standards. Data on the CMT is reported at the school and district levels, but not the state levels, by race, gender, free/reduced lunch, IEP or LEP status. This is not done for the CAPT.

Evaluation. No data were available on bias reduction efforts, making complete evaluation impossible. More work probably needs to be done on inclusion. State level reporting by demographic groups should be done on the CMT, and demographic reporting should be done on the CAPT at all levels.

Standard 3: Professional development.

Print materials are provided for professional development to educators. Workshops, handbooks, sample tasks, lessons, sample student work and brochures were provided for professional development.

Evaluation. No data were available on whether Connecticut requires education in assessment for students in pre-teaching programs. We cannot tell from the information provided the extent of professional development for in-service teachers, but it appears to involve a variety of kinds of assessment work.

Standard 4: Public education, reporting and parents' rights.

Connecticut publishes a profile reporting school, district and state data. Print materials are available to parents and policymakers to explain the assessments.

Evaluation. FairTest does not have enough information to evaluate the state's performance on this standard.

Standard 5: System review and improvement.

Evaluation. FairTest has no information for evaluating Connecticut's performance on this standard.

Connecticut declined to participate in the survey. This report used two years of CCSSO/NCREL reports, plus CCSSO and AFT reports.

DELAWARE

Summary evaluation.

Because Delaware is launching a new state assessment system and last year administered only a writing exam, it is not scorable. Its approach initially appeared positive, based on SEA plans, its current writing assessment and the stages the state went through to develop the new assessments. However, legislation passed on June 30 requires both norm-referenced and standards-based exams, and the state will implement a high school exit exam. If all this is implemented, what appeared to be a program that would require only modest improvement now may be a program that needs major improvements. The NRT and the graduation test should not be implemented. In all other aspects -- the nature of the standards-based assessments, inclusion and bias reduction, professional development, reporting and review -- it remains to be seen how strong the program will be.

Standard 1: Assessment supports important student learning.

Delaware has content and performance standards and curriculum frameworks in English language arts, math, science and social studies. Additional standards are in development. The standards were used to guide development of the current writing assessment. The new state testing program will be based, at least in large part, on these standards.

In the 1995-97 school years, Delaware administered only a writing assessment in grades 3, 5, 8 and 10, the Writing Assessment Program. Teachers were involved in developing and field testing the assessment and scoring guides and in selecting the final prompts. Students were given time to draft and revise in a 2.5-hour exam. Scoring guides were developed by state teachers, the SEA and the contractor. Anchor papers, which are used to help define the performance levels, were selected from student work by teachers, but scoring was handled by a contractor. Students may request a rescoring.

From 1993-95, Delaware had an interim assessment program which combined norm-referenced multiple-choice items and performance items. This assessment was preparatory to implementation of the Delaware State Testing Program (DSTP) in 1997-98. The new assessment will employ multiple-choice, constructed-response and performance items, in reading and math in grades 3, 5, 8 and 10, starting in 1998; and science and social studies in grades 4, 6, 8 and 11, starting in 1999 and 2000. The writing assessment will be continued. Teachers, administrators, SEA personnel and outside experts are involved in developing the new assessments, but not parents or community organizations.

The purposes of the writing assessment are to improve curriculum and instruction and to report results to the public. The writing assessment is not used for making decisions about students or schools. Consequences of the assessment were investigated by using a teacher questionnaire.

The recently passed legislation requires the SEA to obtain nationally-normed data as well as data based on state standards. The state will either purchase an off-the-shelf NRT or obtain norm-referenced items to include in its other tests that will produce comparative data.

The recently mandated graduation exam will include math and language arts for the class of 2002 and science and social studies for the class of 2005. Other variables may be considered, but the test will be the primary one. Alternative assessments will be allowed, but they must be certified by the SEA.

Evaluation. One critical question will be the quality and implementation of the DSTP. Use of multiple methods -- preferably with only a minor part multiple-choice and the major portions extended-response or performance assessments -- is positive. Cautions should be raised about testing twice in elementary school and possible overtesting in grade 8. The purpose of the writing program is positive, and similar purposes should guide the use of the DSTP. The adoption of an NRT, however, is regressive, as will be implementation of a high school exit test. The state should reverse itself and drop both the NRT and the high school exit test, even though the exit test will allow some options.

Standard 2: Assessments are fair.

The state has a bias review committee to examine the writing prompts for various purposes, including for grade-level appropriateness. The bias review committee can reject or modify items. The committee is selected for balance of gender and racial/ethnic representatives. The writing scores are reported by gender, race/ethnicity and a low-income category.

Approximately 9 percent of tested students have an IEP, and 1 percent are LEP. For the 1996 writing test, 98 percent of the state's students were eligible, and over 90 percent of those eligible took the assessment. The state's intention is to develop new assessments that allow all students to participate.

Evaluation. If the bias review procedures, reporting by demographic categories and inclusion are carried through to the new assessments as they are on the writing assessment, Delaware will have solid performance on this principle, as it already will rely on multiple methods.

Standard 3: Professional development.

The state has no required pre- or in-service assessment training for teachers or administrators. Various courses and trainings are available. For the writing assessment, workshops throughout the state were used to familiarize teachers with the materials. The writing test is scored by a contractor. Use of the interim assessments was considered a form of professional development for the new assessment system, and educator involvement in development of the DSTP is positive. The University of Delaware has conducted surveys of teachers and administrators regarding their professional development needs. The state also uses questionnaires to evaluate the effectiveness of its professional development programs.

Evaluation. More systematic and extensive professional development will be required for prospective and current teachers. The use of surveys and questionnaires is positive, as is reliance on teachers to help develop the new assessments. Both the writing assessment and the new performance assessments should be scored by teachers.

Standard 4: Public education, reporting and parents' rights.

Writing scores are reported in English to students, parents and the schools in three months, and to the public in four months. The state has surveyed to find out what information the public and parents want. The state provides guidance on using results to all but the general public. Examples and scoring guides are available to students and parents as well as educators.

Evaluation. If the state extends its current reporting to the new assessment system, it will be doing a good job. It needs to be sure that public education about the new assessments is very extensive, including examples of student work at various levels of quality, and guidance on interpreting scores. The state should also ensure that its reports and information are understood by the public. The current practice of allowing students to appeal a writing score should be extended to the new assessments, which should also permit parents to review assessments.

Standard 5: System review and improvement.

As the new system is not in place, review and improvement cannot be evaluated. The process of learning from previous assessments and using them as preparation for the DSTP is positive, as is the survey of teachers to determine the impact of the writing assessment. An extensive and regular review process should be instituted that will involve educators, informed members of the public and outside experts, and that will consider alignment to the standards, technical quality, ability to assess critical thinking in each domain, and impact on curriculum and instruction. The state should also review local assessment practices.

Delaware responded to the full FairTest survey. This report also used the CCSSO/NCREL, CCSSO and AFT reports. The state reviewed the draft description and provided last-minute information by telephone.

FLORIDA

Summary evaluation.

While Florida did not respond to the FairTest survey, it appears from other data sources that the state's assessment program needs complete overhaul. The state relies mostly on multiple-choice tests, though a new exam will include constructed-response items in math, has a high-school exit test requirement and mandates that districts use an NRT in two grades and two subjects. The test burden is not very high as only reading, writing and math are assessed. Bias review information was not available, but inclusion efforts need improvement, especially for LEP students. While the state says that professional development is on its agenda, its reported efforts focus only on writing, and thus it appears this area also needs substantial improvement. FairTest has little information on reporting, none pertaining to rights and none about review.

Standard 1: Assessment supports important student learning.

Florida has new content and performance standards in English, math, science, social studies, fine arts, foreign languages and health/physical education, and is finalizing curriculum frameworks based on the standards. It plans to have aligned reading and math assessments beginning in 1998.

The Florida assessment program has included the Grade Ten Assessment Test (GTAT), a custom developed, norm-referenced, multiple-choice test in reading comprehension and mathematics given in grade 10. That test was scheduled to end with the 1996 administration and be replaced in 1997-98 by a standards-based exam, the Florida Comprehensive Assessment Test (FCAT). It will test reading with multiple-choice items at grades 4, 8 and 10; and math at grades 5, 8 and 10, with multiple-choice, gridded-in, and short and extended constructed-response items.

The state also has the High School Competency Test (HSCT), a criterion-referenced, multiple-choice test in math and communications (reading and writing) given starting in grade 11. Students must pass it to receive a standard high school diploma. Student are allowed to take the test up to five times during the eleventh and twelfth grades and as many times as necessary thereafter.

The state also has a writing assessment component, the Florida Writing Assessment Program, which uses responses to SEA provided prompts to assess students in grades 4, 8 and 10. The prompts are used to assess a variety of genres (story, explanatory, persuasive). Students receive 45 minutes to produce a sample on demand with no revisions permitted. Scoring is by a commercial company using a state rubric.

The state requires districts to administer NRTs to students in grades 4 and 8 and to report the data in reading and mathematics to the state Department of Education. For these tests, each district must select a score below which students will receive remediation.

All tests are intended to be used for improvement of instruction. The results of the writing component are used for program evaluation. Results of the HSCT and the writing assessment are used for accountability, and the HSCT as part of identifying low-performing schools. Schools that perform poorly on tests over a three-year period could face intervention by the state.

Evaluation. The strong reliance on multiple-choice items -- which will continue even with new assessments, the graduation test, and the mandated NRT are all fundamental aspects of the state's program that need to be changed. The goals should be minor reliance on multiple-choice items, no high school exit exam, and no mandated NRT. The time allotted for writing is insufficient, and writing to a prompt is itself a limited means of assessing writing, though allowing different genres is a positive step.

Standard 2: Assessments are fair.

Print materials are provided to students to explain the tests. Data are released by race and gender at the school and state levels. Some IEP and LEP students are not required to be tested, but a student must pass the HSCT to attain a regular high school diploma. A fairly wide range of accommodations are available on the state-made tests for students with IEPs, but few are available for students with LEP. Results of those tested are not included in regular reports, nor are separate group reports issued.

Evaluation. No information on bias review was available. Data also should be released by SES and IEP and LEP status. IEP and LEP students also should be included in regular reports. Inclusion in assessment needs to be strengthened for IEP students and greatly improved for LEP students.

Standard 3: Professional development.

The state reported to the CCSSO that it recognizes the need for widespread professional development. Print and video materials for professional development are available to educators. A computer-based staff development program to introduce teachers to the scoring procedures for the writing test is being field tested.

Evaluation. Though the SEA apparently recognizes the need, widespread professional development does not seem to be available in Florida. The computer-based program could be promising, but writing is only one subject area. No data were available on pre-professional requirements or whether the state evaluated teacher competence in assessment.

Standard 4: Public education, reporting and parents' rights.

The state provides print materials to policymakers and the public for educational and information purposes.

Evaluation. No information on parent or student rights or more detailed information about public education on assessment issues was available.

Standard 5: System review and improvement.

No information was available.

Florida declined to participate in the survey. This report used two years of CCSSO/NCREL reports, plus CCSSO and AFT reports.

GEORGIA

Summary evaluation.

Georgia's assessment system needs a complete overhaul. Elements of the state testing program are on hold, but it is not yet certain what will replace them. In any case, the state relies too heavily on multiple-choice, uses NRTs, has a high school exit exam, and tests too frequently. Thus, the whole system, not just parts of it, needs to be redesigned. Reporting of data by sub-populations should be implemented and inclusion should be strengthened. Professional development appears somewhat misdirected toward test interpretation rather than classroom assessment. A more comprehensive review process is strongly needed.

Standard 1: Assessment supports important student learning.

The state mandated Quality Core Curriculum (QCC), which contains the state's standards and goals, is currently being revised. The current version is no longer required. QCC has included standards for English language arts, math, science, social studies, various arts, physical education, health, foreign languages, vocational education and various high school elective courses.

The Georgia Kindergarten Assessment Program (GKAP) -- an individually administered, criterion-referenced test with a mix of multiple-choice items, performance testing and a teacher observation element used to assess readiness for first grade -- is also being redesigned. Until the redesign is completed, however, the GKAP remains a requirement. Skills assessed are communication, logical/math, personal/physical and social.

The criterion-referenced, multiple-choice Curriculum-Based Assessments (CBAs) are multiple-choice tests aligned to the QCC. Because the QCC is being revised, the CBAs have been made optional for districts. They were, and will be for those districts continuing to use them, administered to students via matrix sampling. It is available in grades 3, 5 and 8 in math, language arts, reading, science, social studies and health. Districts can choose to use only some of the tests.

The criterion-referenced High School Graduation Test (GHSGT) tests 11th graders in the areas of language arts, math, science, social studies, and writing. Other than a writing sample, it is entirely multiple-choice. All students must pass it in order to receive a diploma.

A writing assessment in grades 3, 5, 8 and 11 uses SEA provided prompts that were developed in collaboration with Georgia educators and a contractor. All students at tested grade levels are assessed with the same prompt. Students are given 75 minutes to produce a writing sample on demand. Scoring is done by Georgia educators with the contractor based on a rubric they jointly developed. This test is still required of all students.

All the above tests were developed in-state with involvement from state educators.

The norm-referenced, multiple-choice ITBS battery (language arts, science, social studies and math) is given to all students in grades 3, 5 and 8.

All component results are used for the improvement of curriculum and program evaluation except for the GHSGT and the GKAP. The CBA is used to evaluate school system implementation of the QCC. NRTs, GKAP and the writing assessment are used for student diagnosis or placement. The writing assessment, CBAs and NRT results are all used for accountability for schools in the form of school performance reporting. NRT results also are used for school awards or recognition.

Evaluation. Other than a writing sample to a prompt and the kindergarten test, the state relies entirely on multiple-choice tests, including a norm-referenced battery, the high school exit exam and the now-voluntary CBAs. The state should redesign the CBAs to be primarily performance assessment and continue them as a matrix-sampling exam. It should drop use of the NRT. The state should also eliminate the high school exit exam.

Standard 2: Assessments are fair.

A review committee examines items for bias as part of the test-development process for Georgia-made tests. Statistical analysis is also done to detect possible bias in items. Data disaggregated by sub-populations is not reported.

Nearly 10 percent of students tested have an IEP, while fewer than 1 percent are LEP. Some IEP or LEP students are excluded from assessment. Limited accommodations on some assessments are available for IEP. Results for those tested are included in regular reports, except some non-standard administrations.

Evaluation. The bias review committee may need authority to delete or modify items. Sub-population data should be reported. Accommodations and alternatives should be expanded to enhance inclusion.

Standard 3: Professional development.

The state has some pre-service requirements for teachers in assessment. The state has professional development programs in many areas, including assessment. The state recently mandated that teachers in grades 3-12 participate annually in a staff development program on the use of tests to improve students' academic achievement within the instructional program. The topics will include curriculum alignment and disaggregating data by sub-tests. The state has not evaluated teacher competence in assessment. It has not surveyed classroom, school or district assessment practices.

Evaluation. While professional development is being expanded, it seems to focus mostly on the narrow state tests rather than on classroom and performance assessment, which it should do. The state should evaluate teacher competence in assessment.

Standard 4: Public education, reporting and parents' rights.
Print and some video materials are available to students and the public for information purposes. Parents can review assessments after administration under some conditions.

Evaluation. Positively, parental review is possible. Public education may be adequate for the limited format range of the state assessments.

Standard 5: System review and improvement.
A formal review of the state program exists on a limited basis. The need has been recognized for more comprehensive review.

Evaluation. A more comprehensive review, covering technical and consequential issues, is needed. This should include a specific study of the kindergarten test, and studies of the impact on curriculum, instruction, student progress and the ability of the tests to measure critical and complex thinking in the subject areas. As districts may be using the tests for additional purposes, such as grade promotion, the state should also survey district practices and counter any misuse of tests. The state needs to consider how high-quality standards can be assessed in ways that support important student learning.

Georgia responded to the short form of the FairTest survey through a telephone interview, which was followed up by further questions by telephone. The CCSSO/NCREL, CCSSO and AFT reports were used. The state reviewed a descriptive draft.

HAWAII

Summary evaluation.

The state system needs a complete overhaul. It is too early to tell if a proposed new system will accomplish that. The state relies entirely on multiple-choice, mostly uses an NRT, and has a graduation exit test. The NRT is not based on state standards, and the state recognizes that the high school exit exam only partially assesses its own standards. The test burden is only moderately heavy. While the bias review procedures are adequate, options for LEP students are not. Professional development is seriously inadequate. Reporting and public education efforts are limited, but positively the state has surveyed the public to determine what information it wants and whether the reports are understood. The review system is almost non-existent.

Standard 1: Assessment supports important student learning.

Hawaii has only one statewide district. The state has goals, content and performance standards, assessment and curriculum frameworks and student expectations in language arts, math, science, social studies, fine arts, health and fitness, world languages and home and work skills.

Hawaii's assessment program includes an NRT (Stanford 8), used to assess students in grades 3, 6, 8 and 10 in the areas of language arts, mathematics and reading. This test assesses only basic skills and not the rest of the curriculum. Also required is the Hawaii State Test of Essential Competencies (HSTEC), a criterion-referenced, multiple-choice high school exit exam given beginning in grade 10. Students must pass it to graduate. The state also has a voluntary, criterion-referenced, multiple-choice Credit by Examination (CbyE) offered to students in grade 8, in the areas of algebra and foreign languages. Except for the voluntary CbyE, all students are tested.

Results of the NRT and the HSTEC are used for school performance reporting. The NRT is used as one of several indicators of school status and improvement efforts, and for individuals as one part of determining admission to gifted and talented programs. It is intended to guide curriculum and instruction, but no studies have been done on the consequences.

The HSTEC is used in determining receipt of a high school diploma. Students have multiple opportunities to take the test, including two times after leaving school. It is not intended to guide curriculum, but schools may adapt curricula to help ensure students pass the test. HSTEC is designed by SEA staff with outside consultants. Teachers and administrators join in writing items and selecting examples.

Beginning in 1996, the state plans to integrate assessment into a Comprehensive Assessment and Accountability System (CAAS), which is expected to take several years to develop. It will be based on the state standards. Design of an Hawaii Writing Assessment (HWA) is also ongoing. The SEA plans to continue pilot-testing, in collaboration with the Center for Research on Evaluation, Standards and Student Assessment (CRESST), performance-based tests in various areas. The project is contingent on available funds.

Evaluation. The state should drop the NRT, eliminate the high school graduation requirement and shift to a primarily performance-based system, assessing in a few grades based on the state standards. It is not clear whether this will happen with CAAS.

Standard 2: Assessments are fair.

People from the seven islands and diverse ethnic backgrounds participate in bias review on the HSTEC. Items are analyzed pre- and post-test for bias. The bias review committee has authority to delete or replace items. Bias review for the NRT is conducted by the publisher. Practice tests or sample questions are available to students. The state does not report scores by demographic categories.

Twelve percent of students tested are classified as having an IEP and 6 percent of students tested are LEP. Various accommodations are available to students with IEPs, but none for students who are LEP. IEP and LEP students who do not pass the test are not eligible for a regular diploma. A special education diploma is available, as is a certificate of course completion. The state adult education program offers a diploma based on its own assessment.

Evaluation. Bias review procedures are sufficient on the HSTEC. Accommodations on both HSTEC and the NRT are not, particularly for LEP students. Sole reliance on multiple-choice does not meet the fairness principle, nor does the high school exit exam.

Standard 3: Professional development.

The state requires no education in assessment for teachers or administrators, and it offers little training. Descriptive materials and samples are available to educators. The SEA does not survey for teacher competence or needs, nor does it evaluate school or classroom practices.

Evaluation. Professional development is seriously inadequate. Involvement of teachers in writing the HSTEC is positive. Should a new assessment system be introduced that is substantially performance-based, the state will need to provide professional development. The state should also support professional development in classroom-based assessment for pre- and in-service teachers.

Standard 4: Public education, reporting and parents' rights.

Students can be exempted from the NRT at parental request (less than one percent are). Descriptive materials and samples are available to the public. The NRT is reported, in English only, 6 months after the spring administration. HSTEC results are reported in 1-2 months, in English only. On the HSTEC, but not the NRT, the public has been surveyed to find out what information it wants and whether the reporting is understood. On both exams, the SEA provides guidance on the use of results to parents, educators and policymakers.

Evaluation. Reporting and public education appear adequate for the nature of the tests. The survey for the HSTEC is positive. Should the assessment system change positively, then substantial public education will be needed.

Standard 5: System review and improvement.

The SEA does not review the state system, but does conduct studies on the HSTEC which are used in revising that test.

Evaluation. Regular review of the system is needed. This review should include the consequences of the tests for curriculum instruction and graduation and whether the exams adequately assess the standards.

Hawaii responded to the full FairTest survey. This report also used the CCSSO/NCREL, CCSSO and AFT reports. The state reviewed a descriptive draft.

IDAHO

Summary evaluation.

Idaho's assessment system needs many major improvements. It is far too reliant on multiple-choice NRTs and tests too often. The state should stop using the NRTs and instead continue on the path it is only now starting: developing constructed-response, criterion-referenced assessment in a limited number of grades. It also will need to substantially strengthen its fairness efforts and professional development. Positively, the state does not attach high stakes to its tests, is developing performance exams, and may have a start on a good review system.

Standard 1: Assessment supports important student learning.

In 1995, the Idaho SEA put on hold its previous frameworks and began development of its *Skill Based Curriculum Guides* in math, science, music, art, social studies, language arts (reading, writing, language, spelling), health and physical education. These include sample methods that districts and schools can use to test at each grade level for each skill in each content standard. Exit standards are being developed.

The assessment program includes multiple-choice NRTs in grades 3-8 (ITBS) and 9-11 (TAP) in the areas of language arts, math, reading, science and social studies. The SEA says the ITBS is aligned with the new frameworks as these frameworks "were designed to incorporate the information measured in our state assessments using the (ITBS)." The state assesses writing in grades 4, 8 and 11 using responses to SEA provided prompts. Students in grades 8 and 11 receive 90 minutes to produce one writing sample on demand, and students in grade 4 receive 60 minutes. They are scored by teachers in the state with training provided by the SEA.

The state has constructed-response assessments in math for grades 4 and 8 with scoring guides developed by the state. In the next two years, the SEA plans to develop assessments in science and social studies and further refine the math assessment. These and the writing assessments will be revised or developed to match the standards. They are intended to guide curriculum and instruction.

Results of assessments are used for curriculum improvement and school performance reporting. There are no high-stakes consequences from results for schools or individuals at the state level.

Evaluation. Despite incorporation of ITBS-based data in the standards, the ITBS, as a multiple-choice test of basic skills, is not an adequate means of assessing all areas of a domain, meaning the assessments are not likely to be adequately aligned to the standards. Idaho is far too reliant on multiple-choice NRTs, and it tests in too many grades. The new assessments should be a substantial improvement, since they are based on the standards, criterion-referenced and include more constructed-response items. The primary step the state should take is to eliminate the NRT, or use it to sample in a few grades at most, while developing further its own assessments based on the standards.

Standard 2: Assessments are fair.

The state does not have a bias review committee, but it does "attempt to be sensitive to gender and cultural backgrounds in developing prompts and questions." Development of the math assessment has attempted to take into account different learning styles, including giving students opportunity to choose which 4 of 5 items to answer. The items also have different ways to solve them. Reports do not include data by demographic categories.

On the NRT and the math and writing assessments, some students with IEP and LEP are not tested, and limited accommodations are available for both groups. The results of those tested are included in regular reports.

The state does not provide test practice materials, but former writing prompts are released each year, and ITBS provides some materials.

Evaluation. A stronger bias review committee is in order, particularly as the state develops more of its own assessments. It will need to make a stronger effort to include IEP and LEP students. The approach toward different learning styles in the math assessment appears notably positive.

Standard 3: Professional development.

Idaho has no pre-service requirements in assessment, and does not evaluate teacher competence in assessment or survey teachers regarding professional development needs. Print materials for professional development are available to educators. It provides pre- and post-testing workshops for ITBS and TAP at six sites each fall, and offers writing and math workshops at state meetings and fall workshops around the state.

Evaluation. Substantially more professional development in assessment that meets instructional as well as accountability needs is necessary. Positively, teachers score the writing assessments and also are involved in developing and scoring the constructed-response exams.

Standard 4: Public education, reporting and parents' rights.

Print materials for explanatory purposes are available to educators, parents and policymakers in English. The state has not surveyed parents or the public to determine what information they want or whether they understand current reports.

Evaluation. More extensive public education will be necessary if Idaho significantly alters its assessment system.

Standard 5: System review and improvement.

The SEA reviews the assessment system, and the legislature is reviewing it in 1997. Review includes studying the impact of assessment on curriculum and instruction. Studies have shown improvements in student writing attributed to the assessment. The new performance assessments do not yet have technical studies and have not been reviewed for alignment with standards or for how well they assess critical thinking. The SEA has not surveyed classroom, school or district assessment practices.

Evaluation. The review process, by including a study of the impact of assessment, is positive. Other than this promising provision, we have too little information on other aspects of the review process to comment further.

Idaho responded to the short form of the FairTest survey. This report also used the CCSSO/NCREL, CCSSO and AFT reports. The state responded to a draft description.

ILLINOIS

Summary evaluation.

This program needs some significant improvements, primarily by implementing assessments that use a variety of methods rather than near-exclusive reliance on multiple-choice, by ensuring that the mandated LEA assessments are of high quality and used properly, and by extensive professional development. Positively, the state does well on many areas of equity and public involvement and reporting. The review process appears to need significant strengthening in several important areas.

Standard 1: Assessment supports important student learning.

Illinois has *State Goals for Learning* and is developing more detailed content standards in math, language arts, science, social studies, music, art, health and physical education. Thirty-four state goals developed by teachers, consultant experts, state staff and other educators are currently under review. State content standards are scheduled for Board adoption by fall 1997.

The Illinois Goal Assessment Program (IGAP) tests students in grades 3, 6, 8 and 10 in the areas of reading, mathematics and writing. Students in grades 4, 7 and 11 are tested in the areas of science and social science. Tests are multiple-choice (some with more than one correct answer), norm- and criterion-referenced, plus a writing sample. IGAP items are aligned to specific elements of the current state goals. All students are tested and all students see the same items.

For the writing assessment, responses to SEA provided prompts are used. Students are given 40 minutes for each prompt. They are scored commercially.

In addition to the IGAP, each school in the state must implement an assessment system to measure student achievement in the 34 state goals.

IGAP is intended to guide curriculum and instruction. The Advisory Committee reports that teachers and administrators have used IGAP in this manner. Advisory committees review the developmental appropriateness of tests. The state claims, "Most of the items are designed to have students use critical thinking." Test items are pre-tested and evaluated.

Results of assessment are used for school performance reporting (school report card) and accreditation. Consequences for schools include possible exemption from regulations, probation, funding loss, accreditation loss, takeover or dissolution; however, test results are not the sole criterion for these determinations.

Changes in the state assessment for 1998-99 are under consideration, but decisions have not been made. Performance-based items are under consideration. The state will conduct limited sample testing in art and health in selected grades. The state also plans to implement a high school state assessment for awards of excellence in 1999-2000.

Evaluation. The testing burden is only a bit more than it should be (four grades with same subjects tested instead of three grades) and is reasonably distributed. Expecting LEAs to base their assessments on the state goals is reasonable, but it appears that the states does not now ensure that these assessments are of high quality and have a positive impact. The state does not base high-stakes decisions on a single test for students or schools. Negatively, the state relies entirely on multiple-choice items except for the writing sample (which, at only 40 minutes, is too short) and attempts to combine normative and criterion-referenced data in one exam. Despite state claims, critical thinking cannot be assessed adequately through such heavy reliance on multiple-choice. Using such narrow-format exams to guide curriculum and instruction is also a problem. Of the proposed changes, including performance-based items is a good idea, while awards of excellence based solely on test scores is not.

Standard 2: Assessments are fair.

The state has a bias review committee that includes representatives from major racial/ethnic groups. Committee members are first nominated by either school authorities or special interest groups, then complete bias review training, and finally are selected by assessment staff. The bias review committee can delete or modify items based on data collected. Reported test score data do not include demographic information.

Ten percent of tested elementary school students have an IEP, and 3 percent of tested students are identified as LEP. A statewide assessment of English proficiency in reading and writing in grades 3 through 11 for those students in bilingual education programs who are exempt from IGAP (three years or less in an ESL or bilingual program) was implemented in March 1997. There are no alternative assessments to the IGAP, but valid accommodations for IEP or LEP are allowed. Both groups are included in regular reports.

Evaluation. The bias review committee is strong. Accommodations seem reasonable, but alternative or native-language assessments are needed since the state has many LEP students. Heavy reliance on multiple-choice hinders equity by not meeting needs of diverse learning styles or cultural backgrounds. Reporting should include data by demographic groups.

Standard 3: Professional development.

The state requires no specific professional training in assessment. It does offer trainings on performance assessment and on the IGAP through regional centers. The state does not evaluate teacher competence or their needs in assessment.

IGAP makes printed material -- including descriptions of assessment methods, samples and scoring guides -- available to students, teachers, parents, community and policymakers. Videos are available to students and teachers.

Evaluation. The state should create systemic professional development in assessment for pre- and in-service teachers. It also should evaluate teacher competence and assess teacher needs. Teachers should score the writing samples.

Standard 4: Public education, reporting and parents' rights.

Individual results are available to students and parents in 7 months, while group reports are available to schools and the public in 4.5 months. Results are reported in both English and Spanish, but not Asian languages. The state has surveyed to find out what information the public and parents want. The SEA also provides guidance on the use of results to all but the students.

Evaluation. Public involvement in education is generally positive, including providing reports in Spanish, surveying the public and providing guidance to public on test score interpretation. Negatively, parents cannot review assessments.

Standard 5: System review and improvement.

All stakeholders except community groups are involved in "continuous" evaluations of the state assessments, and all except students and community groups are involved in assessment design and writing, and bias review.

Evaluation. Positively, the IGAP has relatively strong public and external involvement in evaluating the test, and validation studies do include studies of alignment with goals and the impact of the assessment. A more critical, multi-faceted evaluation of the ability of the tests to assess cognitively complex work and critical thinking should be undertaken, as should a study of the impact of assessment on curriculum and instruction. The state also should evaluate the quality of the mandated local assessments; it used to do some of this, but no longer.

Illinois responded to the full FairTest survey. This report also used CCSSO/NCREL, CCSSO and AFT reports. The state responded to a draft description and to follow-up questions.

INDIANA

Summary evaluation.

Indiana's program needs many major improvements, perhaps a complete overhaul, including: changing the assessment used to a fully criterion-referenced rather than norm-referenced assessments that will fully match the standards; shifting from multiple-choice to mostly performance exams; eliminating the incoming high school graduation exit exam requirement; providing more extensive professional development and involvement of educators in assessment construction and scoring; extending public information and education; and conducting a more thorough and regular review process.

Standard 1: Assessment supports important student learning.

Indiana has content standards in English language arts, math, science, social studies, foreign languages, fine arts, health and physical education. Performance standards are under development.

The state administers assessments (ISTEP+) in grades 3, 6, 8 and 10 in language arts and math. ISTEP+ is currently a customized off-the-shelf, norm-referenced test (CTBS/4 Survey Edition) with criterion-referenced items built in, which uses both multiple-choice and open-ended items. ISTEP+ is aligned to the standards by content mapping via expert agreement. The verbal communication skills element of the state standards is not assessed.

A commercial testing firm is used in the design and development of tests as well as the scoring and reporting of results, with input from SEA personnel and teachers.

Results of statewide assessment are used for student diagnosis, funding for remediation, curriculum improvement, public reporting and program evaluation. Test results are part of determining school monetary awards, probation, accreditation or takeover. In the fall of 1997, a graduation exit exam will be added.

Evaluation. Positively, Indiana includes some open-ended items, though multiple-choice items remain dominant. Combining criterion-referenced items with an NRT is a questionable practice in terms of assessing to standards, as such tests typically are adapted from an NRT rather than from a CRT. It is also questionable whether the test fully assesses the standards, as it is a mostly multiple-choice test. The test burden is only a bit more than it should be (4 grades with same subjects tested instead of 3 grades). Unfortunately, a high school exam test requirement is being phased in. The state should shift the balance to predominantly constructed-response items, should shift to a criterion-referenced instrument that fully matches the standards, and should not implement the high school exit exam.

Standard 2: Assessments are fair.

The state has a multicultural "sensitivity review committee." Items are pre-tested and analyzed for bias. The committee makes recommendations to the Education Department, which makes the final determination. Students appear to be given adequate opportunity to learn about the assessment. Demographic data are not presented.

Eight percent of students tested have an IEP. IEP or LEP students may be exempted from assessment. Various accommodations are allowed for IEP. IEP students are excluded from regular reports.

Test information, including samples, examples and scoring guides, is provided to teachers, administrators, parents and policymakers. A short practice exam is available to students for format familiarity.

Evaluation. The bias review process seems adequate. Accommodations or adaptations for IEP appear to be adequate, but are not adequate for LEP. The pending use of the high school exit test runs counter to the *Principles and Indicators*.

Standard 3: Professional development.

The state has no required pre-service or in-service training in assessment. It offers training in classroom and performance assessment, and workshops on the ISTEP+. Through specific program surveys, the state gathers information on teacher professional development needs in assessment.

Evaluation. Positively, the state offers trainings, but they appear not to be systematically available to all teachers. The state actively gathers information about teachers' education needs in this area, but has not evaluated classroom, school or district practices. Teachers do not score any assessments. Expanded and more systematic professional development is recommended, along with gathering more information.

Standard 4: Public education, reporting and parents' rights.

Print materials are available to educators, parents and policymakers about the ISTEP+. The test is secure, so parents are not able to review it. Test results are reported, in English only, in 4 months after administration to students, parents, schools and the public.

Evaluation. In general, Indiana does not meet this standard well. Some states with commercial tests do allow parents to review the exams under secure conditions. Public information is insufficient, though reporting is timely for a large-scale assessment.

Standard 5: System review and improvement.

The state itself has not formally reviewed the state assessment program, nor does it evaluate or survey district, school or classroom assessment. The ISTEP+ is intended to guide curriculum and instruction, but the SEA has no studies of the consequences using this assessment for this purpose. It has been reviewed for how well it assesses critical thinking, and technical studies have been done. No plans for changes exist.

Evaluation. The review process contains a few of the important elements, but leaves most out. A more comprehensive, independent review process, including a focus on the instructional impact of the ISTEP+ and a study of the ability of the exam to assess cognitively complex work, is warranted.

Indiana responded to the full FairTest survey. This report also relied on CCSSO/NCREL, CCSSO and AFT reports. The state responded to a draft description.

IOWA

Summary evaluation.

Iowa has no state test and does not mandate tests for districts, but it has a voluntary program for districts, approximately 99 percent of whom use one of the Iowa Testing Program's NRTs (particularly the ITBS), as well as other assessments. A district-based approach is reasonable, but the state should discourage extensive use of the Iowa NRTs and emphasize use of performance assessments such as the New Standards assessments the state is helping districts learn to use. The state should also provide guidance on fairness, public education and district self-evaluation. It does offer professional development that appears solid but could be expanded. It might be advisable for the state to review and evaluate district practices to offer support and guidance for improvement.

Standard 1: Assessment supports important student learning.

While Iowa has four broad state goals, which were developed with public participation, the state has adopted the approach of helping districts improve the capacity to improve themselves. Therefore, the state works with districts to help them develop, "through informed dialogue with its community, a clear set of learning expectations... and standards for student performance." The state will provide model standards from professional organizations and other states that districts can adopt or use in developing their own.

Iowa does not have a state test. Instead it has a voluntary testing program in which about 99 percent of the districts use the norm-referenced, multiple-choice Iowa Tests of Basic Skills (ITBS) and the Iowa Tests of Educational Development.

The state expects that districts will use a variety of assessment methods in determining student progress, not just the Iowas. The state has been involved with New Standards, which has assisted local districts in developing alternative assessments. In addition, the state is in the process of identifying multiple assessments to meet the requirements of the federal Title I program of the Improving America's Schools Act.

Evaluation. The approach of a state providing guidance to districts is acceptable. The state should, however, discourage major reliance on NRTs.

Standard 2: Assessments are fair.

Any bias review would be either conducted locally or by the maker of the standardized test used. Whether to include IEP or LEP students is also a local determination.

Evaluation. It would be appropriate for the state to issue guidelines to districts to ensure bias reduction techniques are used, to maximize inclusion through accommodations and alternative assessments for IEP and LEP students, and to guard against use of one test for high-stakes decisions.

Standard 3: Professional development.

Professional development does exist to help teachers learn to select, develop or use new forms of assessment in classrooms. Iowa has funded a State Assessment Center "to promote research and development of local assessments." The primary use of the funds has been to support involvement with New Standards. Funding has been used to train teachers to use portfolios in language arts and math. Recently training has started in the areas of science and applied learning. In addition, each district receives funding for staff development under the state's Educational Excellence Program.

Evaluation.

The focus on improving local and classroom assessments, particularly portfolios, is positive. The state should survey teachers to see if the professional development meets educators' needs for assistance.

Standard 4: Public education, reporting and parents' rights.

Districts are required to report to the community and state about their progress in reaching their achievement goals.

Evaluation. Guidelines in reporting and educating the public, particularly in non-traditional assessments, would be useful. Progress in reaching goals should not be reduced to scores on the ITBS or similar tests.

Standard 5: System review and improvement.

Iowa has no state system to review.

Evaluation. The state should review and report on district practices to provide helpful feedback, or it should ensure that districts evaluate their assessment systems, including technical and consequential reviews. These should include the impact of the Iowas on curriculum and instruction and the ability of assessments to evaluate critical thinking and understanding of cognitively complex material.

Iowa responded to a draft description based on the CCSSO/NCREL report.

KANSAS

Summary evaluation.

Kansas is introducing a new assessment system that probably will need significant improvement. Though constructed-response items have been included, the proportion of multiple-choice remains substantially too high, particularly in reading (90 percent); and writing relies solely on a single writing sample. Portfolios and extended or performance tasks should be included or strengthened. Positively, no NRT is used and the state does not have high-stakes tests. Bias review appears to be solid and inclusion is well on the way to meeting new IDEA requirements. Reporting needs to include LEP and IEP students, and also needs to provide disaggregated data by race, gender and SES. Professional development appears strong, as does public education. Review procedures also appear to be fairly strong.

Standard 1: Assessment supports important student learning.

The state assessments are based on state goals, content and performance standards, and curriculum frameworks, in communications, math, science and social studies. The state recognizes that some areas within standards are not tested.

Kansas has shifted to a format with more emphasis on performance-based assessment of higher order skills, replacing basic skills testing. However, all or part of some standards are not assessed, and extended-response or performance assessment remains a minority of the score for all subjects except writing. The assessments have not yet been analyzed for how well they elicit higher order, cognitively complex thinking.

Kansas tests, or will soon test, at three grade points per subject -- elementary school (grades 3, 4 or 5), middle school (6 or 7), and high school (10 or 11) -- in writing, math, reading, science, social studies and civics. Methods include criterion-referenced multiple-choice, short or extended constructed-response, and individual or group performance assessment. All students in public schools and accredited private schools in a designated grade are tested. Multiple forms are used.

The weight of different methods varies by subject. For example, in math extended response takes 50 percent of the time and earns 25 percent of the score, while multiple-choice takes 35 percent of the time but counts for 50 percent of the score, and short-response takes the remainder. In reading, multiple-choice takes 90 percent of the time and score. Writing consists entirely of a writing sample. In science and social studies, 50 percent of the score is from multiple-choice, 25 percent from extended response.

Assessment development, scoring, analysis and reporting is done by a contractor, the Center for Educational Testing and Evaluation at the University of Kansas. Teachers, administrators, SEA personnel and outside experts were involved in developing the assessments. Community representatives were added for bias review. Students were not involved.

The purposes of the state assessments are instructional improvement, program evaluation, school accreditation and public reporting. There are no high stakes for individuals or schools. Reports to schools emphasize program improvement rather than individual scores. Improvement on the exams is one component of information used for school accreditation reviews.

Evaluation. The new Kansas assessments are a major improvement over the previous basic skills program, but still rely too heavily on multiple-choice and short-answer items, particularly in reading at 90 percent multiple-choice. Writing to a sample provides a limited perspective on writing. Portfolios and more extended-response or performance items should be included or expanded. Positively, the state does not use NRTs or high-stakes tests and keeps the testing burden relatively light.

Standard 2: Assessments are fair.

Kansas views the variety of assessment methods, use of calculators and word processors and allowance of accommodations as means of responding to a variety of learning styles. "Logical review" and empirical analysis are applied to assessment items. The bias review committee recommends changes to the KSDE, but it does not have authority to delete or modify items. There are no demographic requirements for participation on the bias review committee, but the committee has substantial racial/ethnic diversity. Test results do not include reports by demographic categories. The SEA provides explanatory materials and examples to students, but not practice materials.

Over 10 percent of Kansas students have an IEP, and under 2 percent have LEP. Students in either category can be exempted from the state assessment by teacher or administrator recommendation. About 60 percent of IEP and 80 percent of LEP are tested. LEP students can take the assessment with an accommodation. The SEA has established a task force to develop an alternate assessment, accommodations and ethical practices document. Accommodations that are adopted will be available to all students, not just those with disabilities. LEP and IEP are excluded from regular reports and no separate reports are issued.

Evaluation. Giving more authority to the bias review panel might be warranted. Reporting needs to be substantially strengthened to include IEP and LEP in regular reports and to provide disaggregated data by race, gender and SES. Inclusion in the assessments appears to be equal or superior to most states, but still inadequate. However, the task force should lay the groundwork to enable the state to meet the new IDEA requirements for inclusion of all IEP students in assessment.

Standard 3: Professional development.

Kansas has substantial requirements for pre-service teachers and administrators, including learning about traditional standardized tests, various classroom and performance assessment techniques, and the state assessment program. The state offers classroom and performance assessment trainings to in-service teachers and administrators, and a minimum of 10 workshops per subject per year on the state assessments. The state does not evaluate teacher competence in assessment, but has evaluated school assessment practices as part of school accreditation. It does survey teachers and administrators about their assessment needs as part of the state assessment program, and surveys and evaluations accompany the state's trainings.

Evaluation. Kansas appears to have relatively strong professional development for assessment, from pre-service training to the programs for in-service teachers. The inclusion of education in classroom assessment is very positive. The surveys and evaluations of school practices are solid. The accreditation process should be used to ensure teacher competence and high-quality school practices.

Standard 4: Public education, reporting and parents' rights.

The state provides descriptions of assessment methods, samples of assessments, scoring guides and examples of work to students, teachers and administrators. Samples or assessments with examples of work are provided to the community.

The state has surveyed the public as to the content sought in reports and whether the reports are understandable. Public reports are available only in English.

There are no formal procedures for complaints or challenges. Parents can review the assessment at any time but may not copy it.

Evaluation. Public education seems fairly strong, as does the right of parents to review the tests. The state should continue its surveys to ensure the public is adequately informed as the new assessment system is introduced.

Standard 5: System review and improvement.

The state assessment system is evaluated annually. Surveys and interviews of teachers, administrators, parents and community organizations are complemented by regular feedback from education department staff and a review of standards by outside experts. Analysis of alignment and the impact on curriculum and instruction are conducted, as are reviews of school assessment practices. Technical studies have been done.

Surveys have shown that the new assessments are leading to changed instructional practice and improved student learning. Surveys show students are more positive toward the new assessments, particularly in writing and the science and social studies performance assessments.

Evaluation. Because the assessment system is new, the evaluation process is in its early stages, but the initial steps are sound, combining analysis of the standards and assessments with survey data. The state will need to ensure that apparent improvements in learning are not simply inflated scores on the tests. The state should also ensure that district practices are compatible with helping students attain high standards.

Kansas responded to the full FairTest survey. This report also relied on CCSSO/NCREL, CCSSO and AFT reports. The state responded to a draft description.

KENTUCKY

Summary evaluation.

Kentucky's system needs modest to significant improvement. The major questions concern the possible harmful impact of re-introducing an NRT and the as-yet undetermined weight to be given to the reintroduced multiple-choice items on the state assessments. Very positively, the preponderance of time on assessments is devoted to the constructed-response items and the state uses portfolios. The state testing burden is not heavy, but the NRT substantially increases that burden (though the NRTs are not administered in grades assessed with the Kentucky Instructional Results Information System, or KIRIS). Appropriately, the NRT is not used for accountability, which means it probably will have little harmful impact on curriculum and instruction. The stakes on KIRIS are high for schools and staff, which has been a source of controversy, and they may well be too high. Because KIRIS, with its new forms of assessment, has had some serious technical problems, the state is under some pressure to revert to traditional tests. Given a choice between lowering stakes and reverting to traditional tests, the state should lower the stakes. The state also should carefully monitor districts for possible misuse of assessment data for grade promotion, graduation or placement. Kentucky does well regarding the other principles: its bias review and inclusion efforts are solid, professional development is supported, public education and reporting are extensive, and the state's reviews are strong and thorough.

Standard 1: Assessment supports important student learning.

Kentucky has academic standards for reading, writing, math, social studies, science, arts and humanities, practical living and vocational studies which are reflected in six learning goals. Academic expectations, a curriculum framework, performance standards and the state assessments are based on the standards.

Kentucky administers KIRIS assessments in reading, math, science, social studies, arts and humanities, practical living/vocational studies and writing, in different grades -- 4 or 5, 7 or 8, and 11.

All students are assessed in KIRIS grade levels. Multiple forms of each test are used. Each subject except writing is assessed with open-response and multiple-choice items. After not using them, the state reintroduced multiple-choice items to the state assessments in 1996. Typically, per content area the multiple-choice section takes .5 - 1 hour, while the open response takes 1.5 - 2 hours. The weight to be given to different parts of the test in the total score has not yet been decided, but the weight of the multiple-choice probably will be under 25 percent. Students are scored according to four levels of performance in each discipline. Open-response items are scored according to state standards.

Writing includes a portfolio assessment, with teachers and students jointly selecting material. It also includes prompts to which students write responses. A math portfolio assessment is in revision and is planned to be used in 1998-99. The state used performance events for several years, but they have been temporarily suspended for further research and will be reinstated later.

At the end of each test booklet is a questionnaire including a selected-response section in which students choose their answers to a variety of questions about their study habits, demographics and home environment. They are also provided an opportunity to make general comments about their education.

The state has connected its assessment to the National Assessment of Educational Progress by including some NAEP items in the KIRIS assessment. In 1996-97, the state also will begin administering the CTBS test in math and reading in grades 3, 6 and 9 to provide national norm-referenced data; it will not be part of the accountability system.

The primary purposes of the KIRIS assessment are curriculum improvement and program evaluation. Rewards for school that improve and sanctions for those that do not are part of Kentucky's reform program. The state uses assessment results from the KIRIS test and portfolios to reward schools and their staff for improvement. Assessment data are a key part of the information used to target schools which are not improving, including sending assistance teams, or in extreme cases potentially taking over or dissolving a school. Districts can, at their discretion, use test results for such things as grade promotion and graduation, but the state applies no high stakes to individual students based on test results.

Assessment design and the writing of items and scoring guides is done by teachers, administrators, specialists from the Kentucky Department of Education, and members of education organizations. Outside experts help in the design of items. The portfolio assessments were developed in-state, with assistance from a private contractor. They are scored at the school level, with periodic audits of samples to check the accuracy of school scores.

Evaluation. Kentucky has a generally positive approach to ensuring its assessments support important learning by emphasizing constructed-response items. The revised portfolios and performance events should complement the shorter-response items that are the heart of KIRIS. If limited, the re-introduction of multiple-choice items should not be a problem. Re-introduction of an NRT is a step backward; at most, it should be used on a sampling basis. Some studies of Kentucky have suggested that the consequences for schools, even ones such as rewards for teachers, are having some negative effects on schools, and these should be monitored and perhaps modified. The opportunity for students to comment on their learning is also positive.

Standard 2: Assessments are fair.

The Kentucky Department of Education (KDE) reports that it considers the special needs and talents of all students when developing items. Bias review involves various categories of people such as parents, community members, and advocacy and business groups. All test items are assessed by a bias review committee which is 60 per cent white/European and 40% from various minority populations. The committee has authority to suggest changes, modifications and deletions of items. KDE content specialists then make changes in items to address concerns of the bias committee. Results are reported at the school level, with a breakout by race and gender, but not SES.

Approximately 8 percent of the state's students have an IEP. Kentucky has very few students with LEP. Fewer than 2 percent of students with an IEP, and those students with LEP who have been in an English-speaking school program for less than two years, are excluded from the standardized exams. Limited accommodations exist for the regular assessments. Alternative portfolios are designed to meet individual student abilities. These portfolios are evaluated with the same standards as are applied to the general student population, and their results are included in regular reports.

Evaluation. The bias review procedures are solid and the efforts at inclusion are very positive, though more accommodations for IEP and LEP might be warranted. Using a variety of assessment methods should help meet the needs of students with a variety of learning styles. Reporting should include breakouts by SES.

Standard 3: Professional development.

Kentucky's education reform has emphasized professional development. The state allocation for this has risen from $1 per student per year to $23. Each district and school has developed its own professional development plan. Additional training is provided by the state on a voluntary basis.

Pre-service education in classroom observational and assessment techniques and psychometrics are required for teachers and administrators. Other assessment training is available in a variety of ways through the state, and through non-governmental organizations. Evaluation of in-service training is done through evaluation forms at end of sessions. The SEA provides information about its assessments (methods, samples, scoring guides and examples of work) to teachers and administrators. A monthly publication includes continuing information about assessment, and television programs on special channels are used as an instruction vehicle for teachers. Teachers' competence in assessment is not evaluated, but the state has surveyed assessment practices at the district and school levels.

Evaluation. Kentucky has put more focus on professional development, including training in assessment, than most states. Studies of district and school level assessment practices should include some review of teacher assessment practices.

Standard 4: Public education, reporting, and parents' rights.

Kentucky provides information about its assessments (methods, samples, scoring guides, and examples of work) to students, parents and the community. Released items from previous years are part of this process.

Test results are reported to schools and the public within six months of student testing. Schools, in turn, notify students and parents. Reports are made only in English. The state has conducted surveys to find what information parents and the public want and whether the reports are understandable. State reports include guidance on the use of test results. Items are "secure," but parents can be shown items, on request, after administration.

Evaluation. Public education and reporting, including the use of surveys, appear to be solid and comprehensive.

Standard 5: System review and improvement.

The KDE and a private contractor evaluated the alignment of the tests with Kentucky's standards, and included parents, teachers and business leaders in the review. They concluded there are no major curriculum areas that are not tested, but the tests have not been evaluated to determine how well they assess the cognitive complexity of the subject. An annual technical report is completed, including reliability and validity studies. The KIRIS Writing Advisory Committee has reviewed the portfolio assessment. The SEA is aware that tests can have both positive and negative consequences for curriculum and instruction, and has contracted for an impact study.

Kentucky's may be the most studied state assessment system ever, with many kinds of independent evaluations. These studies, together with the state's evaluation of the progress of KIRIS, have led to ongoing changes in KIRIS. The core of the program, primary reliance on the open-response items, has remained constant.

Evaluation. Kentucky has been extensively and thoroughly evaluated and the evaluations continue. It would be important to study how well the assessments measure cognitive complexity and critical thinking as designing such assessments is difficult. The impact study should be used to further consider the accountability system and its positive and negative impact on schools, including both student learning and school climate.

Kentucky responded to the full FairTest survey. For this report, we also used CCSSO/NCREL, CCSSO, and AFT reports. The state responded to a draft description and to subsequent questions.

LOUISIANA

Summary evaluation.

Even with proposed changes that will improve the state's assessment program, the program will need a complete overhaul. New assessments will be a mix of multiple-choice and constructed-response, which will be a positive change. The state should drop its NRTs, which would also bring the test burden closer to the recommended three grades, should drop the use of a test as a high school exit gate, and should increase the proportion of constructed-response items beyond the planned 40 percent. Professional development and public education and information should both be substantially increased. The review process needs only modest improvements.

Standard 1: Assessment supports important student learning.

Louisiana is developing standards/curriculum frameworks to replace existing *Curriculum Guides*. The new standards will be in math, science, English language arts, social studies, the arts, and foreign languages. Standards were reviewed by teachers before board approval.

The SEA administers a Kindergarten Developmental Readiness Screening Program to all kindergartners. The Louisiana Educational Assessment Program (LEAP) includes criterion-referenced multiple-choice tests in math and language arts (based on the current state curriculum guides) in grades 3, 5 and 7, and a multiple-choice Graduation Exit Examination in English/language arts, writing and math (grade 10), as well as science and social studies (grade 11). The state also administers the CAT NRT in grades 4 and 6.

Writing on the high school exit test is also assessed by response to a state-developed prompt. Students are initially given 70 minutes to respond, but are permitted to take more time if necessary. The writing is scored by a commercial company using a scoring guide developed by an advisory task force under state guidance.

The kindergarten assessment uses two off-the-shelf tests, the Developmental Skills Checklist and the Chicago Early Assessment, that combines a norm-referenced multiple-choice test with a performance assessment. The intent is student diagnosis and improvement of curriculum and instruction.

The criterion-referenced test and the exit exam, with writing samples, are used for student diagnosis or placement, improvement of curriculum and instruction, program evaluation, student promotion (not used as a sole criterion) or graduation, student awards and school reporting. School-level scores are reported publicly. The tests were developed with university, consultant and state teacher participation. Students in state board approved private schools may opt to take the exit exam, and it is optional for home-schooled students. Students are provided with practice exams and can take the test up to 6 times.

The purposes of the achievement NRT are national comparative data, improvement of curriculum and instruction and program evaluation. District-level scores are reported publicly.

Revising LEAP to meet the new standards has already begun. New grade 4 and 8 exams in English language arts and math are expected for 1998-99, to be followed by science and social studies for those grades, and then by a new high school exit test. The intent is for the exams to be about 60 percent multiple-choice and 40 percent constructed-response, including one extended task in each exam. Assessment prototypes for classroom use are also being developed, first for math and science and subsequently for other subject areas.

Evaluation. Positively, the state is adopting new standards and developing new assessments to match them, which will employ mixed methods. However, they will still be somewhat too heavily multiple-choice. Negatively, Louisiana's current CRT is not adequate, it will retain the NRT achievement and readiness tests and it has a high school exit exam. The test burden is somewhat high and could be lightened best by dropping the achievement and readiness NRTs.

Standard 2: Assessments are fair.

A bias review committee, including educators and the public, with specified racial/ethnic composition requirements, reviews test items and has the authority to recommend removal or alteration of items. Item statistics are also analyzed. Gender and race data are reported.

About 6 percent of students tested have an IEP and 1 percent of students tested are identified as LEP. Both IEP and LEP students may be excluded or receive accommodations from the CRT, NRT and high school exit tests. For the CRT, scores of those tested are included in regular reports and released separately. For the NRT, IEP students with test modification(s) are excluded from regular reports, while some LEP students are included and some excluded. No separate reports are released. IEP and LEP have the same high school graduation requirements as regular students.

Evaluation. Bias review procedures are solid, and efforts are made at accommodations for IEP and LEP. The graduation requirement is a serious problem, as is the over-reliance on multiple-choice methods.

Standard 3: Professional development.

The state has no requirements for teacher professional knowledge in assessment, either pre- or in-service. It offers information on traditional standardized tests and the state assessment programs, and classroom and performance assessment education is available through the state for teachers in some specific programs. Models for integrating curriculum and assessment are part of the new state curriculum frameworks, and these have been reviewed by outside experts, parents, educators and business people. Professional development based on the standards is expected. The state has not surveyed district or classroom assessment practices, nor teachers for their needs.

Evaluation. Professional development is not adequate, but the state plans more extensive training based on the new standards and assessments. It remains to be seen whether it will be sufficient and whether the classroom assessment education is adequate. The involvement of teachers along with the SEA and outside experts in writing LEAP items is positive. The state should involve teachers in scoring the new open-ended assessments and the writing samples.

Standard 4: Public education, reporting and parents' rights.

Little public information about state testing is produced. Test scores are reported, in English only, at the state and district levels. Students can appeal their scores on the tests. Parents or students can look at completed assessments by request to the state.

Evaluation. Public education is not adequate. Particularly as new exams are introduced, the SEA should actively inform the public about them. The review and appeal rights are positive.

Standard 5: System review and improvement.

LEAP is reviewed annually by SEA staff and education organizations, but not the public, and there is no independent review. The impact on curriculum, instruction, and graduation rates is studied, as is validity. The validity of the kindergarten test, however, is not studied.

Evaluation. Positively, the review process is done regularly, but should include outside evaluation when the new LEAP is implemented. Ensuring that the LEAP fully assesses to the content standards should be part of the review process. The validity of the kindergarten assessments should be studied, including its impact on curriculum and instruction and on young children.

Louisiana responded to the full FairTest survey. This report also used the CCSSO/NCREL, CCSSO and AFT reports. The state responded to a draft description.

MAINE

Summary evaluation.

Maine's program needs only modest improvements. The state should not shift from sampling to census testing in areas where it now samples, and it should use performance tasks and portfolios. Educator involvement is quite solid, though more systematic and extensive professional development is warranted. While inclusion of IEP and LEP is positive, attention to bias reduction should be strengthened. Reporting and release of old tests is solid, but surveys of parents and the public should be done. The review process is strong.

Standard 1: Assessment supports important student learning.

Maine has standards in English language arts, math, science, social studies, visual and performing arts, foreign languages, career preparation, and health and physical education. A range of stakeholders was involved in developing the standards, and public hearings were held. The specific content standards and performance indicators are now in the rule-making process. The standards will be used for modifying the state's assessments.

The Maine Educational Assessment administers extended-open-response-item exams at grades 4, 8 and 11 in reading, writing, math, science, social studies, and arts and humanities, and in health in grades 4 and 8. No multiple-choice items are used. The state continues to investigate possible future uses of performance assessments and portfolios. "Extended" open-response typically means paragraph-length responses. The tests are loosely timed: a student may be allowed up to 50 percent additional time to complete each section. Average total testing time is estimated at 5-6 hours, spread over several days.

In reading and math, the exams include a set of common items taken by every student at the grade level, enabling individual scores to be obtained. For writing, each student responds to a common grade-level prompt. In the other subjects, the exams are matrix sampled with each student answering the questions on one two-item set out of 12 such sets, producing only school-level scores. The state is considering ending the sampling and administering common items, and hence obtaining individual scores, in science, social studies, and arts and humanities.

The writing samples are scored by teachers on a 6-point scale. The rest of the exams are scored by the contractor, Advanced Systems, using 0-4 point scales, with each item having a unique scoring guide. Scores are reported using state normative scales for all exams. For reading, writing and math, scores are also reported by the percentage of students at each of four performance levels. In the future, normative reporting may be dropped.

At each administration of the MEA, students also respond to one of two forms of a student questionnaire, mostly about instructional practices. Teachers and administrators also answer questionnaires. Demographic and program enrollment data are also obtained.

Items are written by committees of teachers. The state concludes that item quality has been improving as teachers gain experience. The committees address whether at least some items assess complex and critical thinking about content in each subject area.

The assessments are intended for program evaluation and improvement, and for school accountability to the public. Students may obtain rewards or recognition based on test performance. No other consequences or rewards are attached to the results. The state hopes districts and schools analyze student performance on the items to help improve instruction.

Evaluation. Maine's program needs only minor improvements. The approaches of not relying on multiple-choice, testing only a few grades and some by sampling, low stakes, extensive educator involvement, and using questionnaires are all positive. The burden is reasonable. Shifting away from sampling, however, is a step backwards. It also would be better to use a mix of methods by employing more extended response or developing portfolios. Writing in particular should shift from response to fixed prompt to at least choice among prompts or preferably portfolios. The normative reporting should be dropped.

Standard 2: Assessments are fair.

While the state does not have a separate bias review committee, the teacher committees that draft items use sensitivity and bias review guidelines, and the items also are reviewed by MEA staff. Demographic-based data are not reported.

About 6 percent of tested students have an IEP, and about 1 percent are LEP. Both IEP and LEP students may be allowed accommodations for taking the exams. IEP students may, as a last resort, be excluded from the assessment for any or all subjects. LEP and IEP students are held to the same standards as are all other students. LEP students may take tests in other languages, which are translated at the district level, or have exams read in "sheltered English," except for reading. ("Sheltered English" allows the test administrator to read the exam to a student with information provided to clarify the item. For example, a math item based on calculations about "season passes to the Portland Sea Dogs" might be incomprehensible to students who don't know about season passes or the Sea Dogs, but having that knowledge is irrelevant to the math being assessed.) No other form of assessment (e.g., portfolios) is used. A test booklet must be returned for every enrolled student. The scores of all students who complete all the common sections of the exam, including those with accommodations, are included in the state reports.

Evaluation. The bias review approach is probably adequate, presuming there is sufficient knowledge on the item committees to detect possible bias. Statistical review is recommended. The efforts at inclusion are solid, but more will have to be done to meet new IDEA requirements for students with IEPs.

Standard 3: Professional development.

Pre-service teachers are required to become knowledgeable in classroom, performance, and portfolio assessment as well as integrating classroom assessment and instruction. No further education in this field is required of in-service teachers, but the state does offer workshops in these areas, as does the state university system. Administrators and other personnel have similar requirements and options, and also are expected to obtain pre-service knowledge of traditional testing practices and psychometric principles. The state questionnaires periodically ask teachers and principals about professional practices and development needs in assessment, and about existing professional development. The student surveys also include questions about assessment practices.

Evaluation. The pre-service requirements are solid, but the state should provide more systematic professional development for in-service teachers. The questionnaires and surveys are positive. Educator involvement in assessment development is positive, but it would be better to have more assessments scored by teachers.

Standard 4: Public education, reporting and parents' rights.

Common exam items are discontinued from further test use, and then are publicly released, along with examples of student work scored at different levels, to help the public understand the tests. Detailed pamphlets about the state assessments also are available. Public materials make a forceful case, on consequential grounds, for using the all-open-response format and for shifting to reporting by achievement levels not just norms.

Evaluation. The public information effort seems solid. Surveying parents and the public to determine information they want and whether reports are understood is recommended.

Standard 5: System review and improvement.

Assessment review involves teachers, administrators, SEA and university personnel, outside experts, community groups, and (via questionnaire results) students. A program audit is conducted by the legislature every four years. The state clearly intends assessments to produce curriculum and instruction modifications. The reviews consider the impact of assessment on curriculum and instruction. The state assessments also are subject to ongoing review by a technical advisory committee. Studies of the assessments have found reliability of above .9 at the school level and above .8 at the student level.

Evaluation. Maine's evaluation effort is substantially positive, including its regularity, involvement of many stakeholders and use of outside experts, and the studying of the impact of testing on curriculum and instruction. The reliabilities are notably strong for non-multiple-choice and fully adequate for the purposes of the assessment. More study of whether the assessments fully match the standards and measure critical thinking and cognitively complex material should be included.

Maine responded to the full FairTest survey and sent various reports. This report also relied on CCSSO/NCREL, CCSSO and AFT reports. The state replied to a draft description.

MARYLAND

Summary evaluation.

Though Maryland did not respond to the FairTest survey, it appears from other data sources that the state assessment program needs some significant improvements. While it has one of the best performance assessment programs in the nation in the elementary schools, it also requires a high school exit exam and uses a norm-referenced test, though on a very light sampling basis. The current high school exit exam soon may be dropped, but it remains uncertain as to whether new high school exams will be used as a graduation requirement. If the current exit exam is dropped and the new exams are predominantly constructed-response and not used as high-stakes hurdles, then Maryland's system will need only modest improvement. However, we have little information on fairness, professional development and public reporting, and no information on review of the assessment system.

Standard 1: Assessment supports important student learning.

Maryland has content and performance standards along with curriculum frameworks developed by teachers, school system staff and staff from the Maryland Department of Education. Content standards exist in reading, writing, math, science and social studies for grades 3, 5 and 8. High school standards are being developed.

Maryland's assessment program consists of basic skills high school exit tests, elementary school performance exams, and an NRT. The high school exit test has two components, the Maryland Functional Testing Program (MFTP), a criterion-referenced, multiple-choice test in math, reading and citizenship; and the Maryland Writing Test (MWT), which requires two writing samples, narrative and expository. Students taking the MWT have up to a whole day to complete the exam. Scoring is by a commercial company using a state rubric. The exit tests are administered in grades 7-12, with local school systems determining the appropriate grade level for the first administration. A computer-adaptive version is available. It is used for student diagnosis, curriculum and instructional improvement, school accreditation, and school performance reporting.

The Maryland School Performance Assessment Program (MSPAP) is a criterion-referenced performance assessment, based on the state content standards, for students in grades 3, 5, and 8 in the areas of reading, language usage, math, science, social studies, and writing. It is used to assess application of skills and knowledge, higher order thinking, and the integration of knowledge across disciplines. It utilizes short- and extended-response items and individual and group performance tasks. For the writing test, students are given 40 minutes to draft, and after time for reflection, 50 minutes to finish a response to a prompt. Three samples are obtained per student tested. Teachers write the MSPAP performance tasks and score all the exams using state-developed rubrics. Half the items are replaced each year, and old items are available. MSPAP also includes a student survey about instructional practices and the curriculum. MSPAP is used for curriculum and instructional improvement, accountability, school monetary rewards and penalties, probation and takeover, and school, district and state reporting.

An NRT, the CTBS/5, is given every other year to samples of students in grades 2, 4, and 6 in math, reading comprehension and language arts. It is intended for state and district-level reporting.

The state board has recently approved development of 10 high school exams. Whether passing the tests will be required for graduation will be decided at a later date. If so, these exams will replace the MFTP. Rather than graduation, the exams might be used in determining course grades. It is not clear the extent to which the exams will be multiple-choice or, like MSPAP, performance.

A readiness test is being piloted that apparently will be administered to kindergarten students. Its uses were not specified in the information received.

All test component results are intended for curriculum improvement and program evaluation. MFTP results are used for student placement or diagnosis as well as being an exit requirement. Those who pass receive a high school skills guarantee. Both MSPAP and MFTP results are used for school performance reporting and have the possible consequences of funding gain or loss, intervention, probation or accreditation loss.

Evaluation. MSPAP is a very strong program. Low scores have led to some school takeovers, but it remains to be seen whether this is an effective use of test information. The use of the NRT is reasonable, except it should not be administered in grade 2, since standardized, multiple-choice tests are developmentally inappropriate for young children. The current high school exit exam should be dropped, as it may be. The new exams should not become high school exit hurdles and should be primarily performance assessments, akin to MSPAP. The readiness test should perhaps be scrapped; if it is implemented, the state must carefully guard against misuse.

Standard 2: Assessments are fair.

Print materials explaining the assessments are available to students.

For 1994-95, 11.7 percent of students tested were classified with an IEP, and 1.8 percent of students tested were classified as LEP. Moderately extensive accommodations are available to both groups. IEP students may be excluded if their IEPs call for learning outcomes that are different from regular students. LEP students may be excluded only once. To obtain a diploma, performance standards are the same for those groups as for regular students, and tests are available only in English.

IEP and LEP who are tested are included in regular reports, but no separate information is reported. At the state level on MSPAP, race and gender data are broken out.

Evaluation. Bias review information was not given. More effort at inclusion, with appropriate accommodations and alternate assessments, is needed for IEP and LEP students. Reporting needs to provide disaggregated data.

Standard 3: Professional development.

Print materials about the assessments are available to educators for training. Two-day professional development sessions are available for MSPAP.

Evaluation. Too little information is available. We hope the professional development for MSPAP is widely available, but more needs to be done regarding classroom assessment.

Standard 4: Public education, reporting and parents' rights.

Print materials are available for training and explanatory purposes to parents, and for explanatory purposes to policymakers.

Evaluation. Too little information is available for an evaluation. It is not clear whether the state has done a good job of educating the public about MSPAP and reporting its results.

Standard 5: System review and improvement.

No data were given.

Maryland declined to participate in the survey. This report used two years of CCSSO/NCREL reports and the CCSSO standards report.

MASSACHUSETTS

Summary evaluation.

Massachusetts' new state system, as it is being implemented, needs many major improvements. The system will include mixed multiple-choice and constructed-response criterion-referenced exams, based on state standards. It also will continue to use a recently introduced multiple-choice NRT in reading in grade 3. The state will require a high school exit exam and high stakes for schools and districts. The state has piloted programs for schools to develop local portfolios. The planned bias review committee needs to be implemented and given adequate authority. Alternative assessments for students with IEPs are needed. It is positive that the state is developing exams in Spanish, but it remains unclear how the state will assess other LEP students. The state's financial support of teacher professional development is substantial, but some elements of a comprehensive approach are lacking. Further work in community education and information, as well as bolstering parental rights, is needed. The review system is still being planned. The state board has shown a proclivity to simply adopt exams without adequate public discussion as to their consequences for or relevance to the state's reform efforts. Key recommendations are to drop the NRT and the graduation requirement and strongly support local assessment development.

Standard 1: Assessment supports important student learning.

Massachusetts has developed content frameworks in math, English language arts, science and technology, history/social studies, world languages, arts and health. All but history/social studies have been approved by the Board of Education. A new state assessment program is being developed based on these standards.

Currently, the state is in a transitional phase. It is ending one program, the Massachusetts Educational Assessment Program (MEAP), last administered in spring 1996, and is beginning a new one, the Massachusetts Comprehensive Assessment Program (MCAS). At this time, the state testing program has three components.

At grade 3, the state mandates annual administration of the ITBS, an all-multiple-choice NRT, in reading, spelling and vocabulary. For 1996-97, the state also administered to grade 10 the Iowa Test of Educational Development (ITED) complete battery, also an all-multiple-choice NRT, covering aspects of English, math, science and social studies. Districts which already administer a full battery of another NRT in grade 10 can be exempted from the ITED. On both tests, individual, school and district results will be reported. These tests are not based on the frameworks. There are no plans to re-administer the ITED.

The criterion-referenced MCAS exams for grades 4, 8 and 10, to be based on the curriculum frameworks, are now in the item tryout stage in math, and science and technology. They will use multiple-choice and constructed-response items, probably with about half the score to come from each method. In spring 1998, full implementation of MCAS in math, science and technology, English/language arts (probably including writing samples) and history/social science (if standards approval allows) is planned. World language exams are scheduled for 1999-2000. The state will also develop model assessments for local use in the arts and health. The results will be reported at student, district and state levels in Spring 1998.

According to state legislation, passing the grade 10 test will become a graduation requirement at some point. Students will have more than one opportunity to pass the test. MCAS exams, and perhaps the grade 3 NRT and any NRTs used by districts, will be used to identify chronically underperforming schools for intervention purposes. Regulations governing such use are being developed. State law also calls for developing portfolios and other classroom-based assessments. A pilot project has been funded and involves about 60 schools on a voluntary basis.

Evaluation. The state may be rushing too quickly from standards to exams, allowing too little time to develop high-quality items. Use of the ITED in 1997 was confusing and a waste of time and money. The ITBS reading exam should be dropped and districts should be required to develop or adopt adequate classroom-based performance assessments for reading. While it is not finalized that multiple-choice items will comprise half the exams' scores, they should be a smaller proportion. The writing assessments should build in adequate flexibility. In addition to the problem of using single exams for high stakes for students, draft proposals on using exam results for school and district accountability appear seriously flawed. The classroom-based portfolio project is not now part of accountability and reporting. That is appropriate while the portfolio system is being developed, but runs the risk of making the program invisible and jeopardizing funding, as may be occurring. The portfolio/classroom project should have significant state support.

Standard 2: Assessments are fair.

Question tryout results for the new exams will be analyzed for bias by a committee. Assessment development committees have been recruited to reflect the geographic and demographic diversity of the state, partly to take account of the variety of cultural backgrounds of students. To help assess students with different learning styles, students may provide answers through drawings or other paper-and-pencil options on some items.

About 17 percent of the state's students have an IEP, and 4 percent have LEP. For IEP students, the student's team decides whether the test can be taken with or without accommodation or whether an alternative is needed. Students who need alternatives for Iowas or the MCAS will be exempted since no alternative is available. Non-English speaking students who have been in school in the US three years or less, and who are not recommended for regular education the next year, are exempted. Spanish-language exams for MCAS in math and science and technology were tried out in the spring of 1997. Other Spanish-language exams may be developed. To obtain a diploma, all students will have to pass the assessment.

Evaluation. The bias review committee should have been implemented before item tryouts and should have been involved in assessment development. The effort at providing alternative modes of answering for students with different learning styles is commendable; the effort should be studied as part of the system review process. The development of Spanish-language exams is positive, but it is not clear what will be done for LEP students who speak other languages. The state has an unusually large proportion of students with IEPs, so it needs to make a stronger effort at developing alternative assessments. These should be available for the graduation exam, so long as that exam exists. Similarly, exams in Spanish should include the graduation exams. Comparable alternatives should be available in other languages.

Standard 3: Professional development.

Massachusetts has no requirements for pre-service training in assessment for educators, nor has it evaluated teacher competence in assessment. Teachers are asked what their needs are for professional development. The state currently allots $50 per student per year for teacher professional development.

Evaluation. The $50 allotment, provided for in the state's school reform law, is high relative to most states but probably still insufficient to meet the mandates of that law. The absence of requirements for pre-service training, of surveys regarding teachers' needs, and of systematic approaches to preparing teachers for new assessments, undermine the positive intent of the professional development allotment. The portfolio/classroom assessment project is an important area of staff development and needs strong state support and rapid expansion.

Standard 4: Public education, reporting and parents' rights.

"Common items" on the MCAS will be publicly released after test administration. Parents will not be permitted to exempt their children from the tests. Public reports will be released in English only. The public has not been surveyed to determine what sort of assessment information it wants reported.

Evaluation. Release of common items is positive, and the state should ensure that adequate explanation of those items, including scoring guides and sample work on constructed-response items, should be part of the release. Parent and student rights should be strengthened. Reports should be in languages other than English, particularly Spanish. In all, more attention probably needs to be given to this area.

Standard 5: System review and improvement.

The state last formally reviewed its assessment system in 1990. Review plans, including whether to evaluate the impact of the assessments on curriculum, instruction, and school graduation rates, have not been set. The assessments are intended to guide instruction based on the frameworks. Not all aspects of the frameworks will be assessed; the frameworks have noted those aspects that are "best assessed at a classroom level."

Evaluation. A strong review system is essential, but it remains unknown whether Massachusetts will develop one. Because they are new, the exams have not been evaluated to determine whether they are aligned to the standards and whether they assess higher order or critical thinking. Ensuring external as well as internal reviews is also important.

Massachusetts responded to the short form of the FairTest survey and to subsequent telephone questions. This report also relied on CCSSO and AFT reports. The state replied to a draft description.

MICHIGAN

Summary evaluation.

This program needs some significant improvements. The current program uses multiple-choice and some constructed-response items at three grade levels. The writing assessment provides more flexibility than the usual response to a prompt. Stakes are mostly not too high. Professional development has a good foundation through the pre-service requirements and in-service training. Reporting appears mostly adequate. Reviews are limited. Recommended improvements include significant expansion of constructed-response tasks and portfolios, tests in languages other than English, reports in other major languages, and more systematic review of the assessment program and its impact.

Standard 1: Assessment supports important student learning.

Michigan has "essential goals" and a model core curriculum. It is developing detailed curriculum frameworks that will include content standards, benchmarks, instructional vignettes, performance levels and examples of student work in math, English language arts, science, social studies, arts, foreign language and physical education. The state has completed development of assessment frameworks. A new assessment plan will be developed for 2000-2001.

The Michigan Educational Assessment Program (MEAP) assesses students in grades 4 and 7 in math and reading and grades 5 and 8 in writing and science. Social studies tests are planned for 1998-99. The High School Proficiency Test (HSPT) is administered to award state-endorsed diplomas in language arts (reading and writing tests), math and science; students who do not achieve a "proficient" score can earn a local diploma. However, the HSPT is quite long, totalling 11.5 hours. Both MEAP and HSPT employ a combination of multiple-choice and open-ended items, plus a writing assessment. HSPT scores were included in student transcripts, but after substantial controversy, that policy has been suspended.

An Employability Skills Portfolio for grades 7-12 also selects a sample of students at grade 11 for state reporting. It is voluntary for students, school and districts, and it is used for curriculum improvement and as an indicator system for school-to-work.

The HSPT writing test has three parts: 1) critical evaluation of two pieces from the student's portfolio; 2) response to a prompt based on a set of materials that do not involve extensive reading; and 3) response to another, broadly defined prompt (connected to the second part and following time allowed to discuss part 2 in small groups) that allows the student to respond with a genre of choice (e.g., narrative, fiction). The other HSPT tests, each about two hours long, employ multiple-choice and open-response items, with multiple-choice comprising 65 percent of the math score, 89 percent of reading and 72 percent of science.

MEAP science tests include the same elements as the HSPT and also a hands-on science investigation. The score derives about 80 percent from multiple-choice items and is untimed. Stimuli for writing include both written and graphic/cartoon prompts. Students have three hours, including small group and drafting time, to produce one final piece. Great latitude is given regarding genre and focus within the topic. The grades 4 and 7 Essential Skills Reading and Math Tests are multiple-choice and untimed.

Students may receive Certificates of Recognition for sufficiently high scores on MEAP tests. On reading and math, students are notified if their scores are on track toward earning an endorsed diploma. MEAP has no additional consequences for students. MEAP is intended to support and reinforce the voluntary state curriculum. Results of MEAP and/or HSPT are used for curriculum evaluation, school awards/recognition, performance reporting, and school accreditation. Possible high-stakes consequences for schools are warnings, probation, funding loss, accreditation loss, takeover and dissolution; test scores are only one component.

Evaluation. Positively, the state does not use an NRT, but it relies too heavily on multiple-choice items. Properly, it does not use tests as the basis for decisions such as graduation, though even awarding an honors diploma based on a test score is questionable. The test burden is reasonable, except perhaps for the length of the HSPT. Portfolio development is a positive step, though its use and the model of the portfolio are quite limited. Students are given an opportunity to reflect on their learning in the writing assessment, which is a substantially better assessment than the typical response to a prompt. Significant improvement is still needed, particularly the expansion of the proportion of constructed-response items and the use of portfolios.

Standard 2: Assessments are fair.

Bias review for MEAP is based on a committee with broad representation which reviews all items. The committee has authority to delete or modify items. Statistical analyses are also used. Reports at the state level include data disaggregation by race and gender.

About 11 percent of the state's students have an IEP. The state reported under 2 percent and GWU reported 3 percent of the state's students are LEP. All students with an IEP are to be tested by MEAP unless most of their language arts instruction is not in the regular classroom. For accountability measures, results are included. Students with LEP may be excluded if they have not been in the US for two years. Extensive accommodations are available on MEAP and HSPT. For MEAP and HSPT, models of assessment format are available to schools and students, but the state does not encourage practice testing.

Evaluation. The bias review process is acceptable, and accommodations are reasonable for IEP and LEP, though translations for students with LEP should be considered. Students are informed about the assessments. Alternatives or further accommodations for currently untested students with IEPs will be needed to meet IDEA requirements.

Standard 3: Professional development.

For pre-service teachers, an SEA standard requires that they learn to use multiple approaches to assess student abilities. It also requires colleges to meet NCATE standards, which include mandates regarding assessment competence. The state offers professional development opportunities in classroom, performance and portfolio assessment and on state assessment programs to teachers and administrators. Education in traditional tests and psychometrics is not offered by the SEA but is available in the state. Regional school improvement facilitators are asked to identify needed training; assessment will be the focus for professional development for 1997-98. Educators, SEA staff and outside experts are generally involved in writing items and rubrics and selecting examples, while scoring is done by contracted companies.

Evaluation. Michigan does fairly well with professional development, including basic pre-service requirements and fairly extensive training opportunities. The SEA has some formal means of gathering information about teacher professional development needs. Hopefully, the coming year-long focus on assessment will be extensive and systematic for all teachers. Scoring should be done by teachers.

Standard 4: Public education, reporting and parents' rights.

A wide range of educators and non-educators have been involved in assessment design for MEAP and HSPT. A parent guide for HSPT is in production, and models of MEAP are available on the world wide web.

Parents can exempt their children from MEAP or HSPT; 1- 2 percent are exempted. Tests can be reviewed only in the state MEAP office. A few focus groups were used to determine information that parents or the public wanted on the HSPT. The state has not surveyed to determine whether reports are understood.

Results are reported back in 4 months, in English only. The state reports scores on all tests by 3 proficiency levels, 2 for writing. State provides guidance on use of results to all but the general public.

Evaluation. There has been more than usual public involvement in developing the assessment. Public information is being improved. Reports should be in languages in addition to English. Parent rights to exempt their children and review tests are positive.

Standard 5: System review and improvement.

The SEA has not reviewed district or classroom assessment practices, nor has it reviewed the state assessment program in general. No study of the consequences of test use has been done. For HSPT and MEAP, each item is linked to curriculum frameworks and the balance of items is reviewed to ensure alignment. However, not all important areas are assessed on all tests. The intent in both MEAP and HSPT is to assess application, problem-solving and understanding; for example, the HSPT math test has no purely computational items. However, no formal evaluation of the cognitive complexity of the assessments has been conducted. A technical advisory committee reviews HSPT. Technical studies are underway; some are near completion.

Evaluation. Review and evaluation need strengthening. Review and validity studies have not been done (except for the employability portfolio), nor has the ability of the assessment to assess critical thinking skills been evaluated. Forthcoming technical studies may inform program improvement.

Michigan responded to sections 1-3 of the full FairTest survey and sent extensive documents as a response to section 4. This report also relied on CCSSO/NCREL, CCSSO and AFT reports. The state replied to a draft description.

MINNESOTA

Summary evaluation.

Minnesota's current program needs many major improvements. The legislature recently adopted some modifications to the current program, but these do not fundamentally alter the state's program.

The Basic Standards Tests (BST) are the only tests currently in place. They are multiple-choice with a writing sample and include a high school graduation test requirement. The high school graduation requirement should be dropped and the BST either dropped or substantially modified, but the legislature affirmed both and allowed the SEA the option of a norm-referenced test.

The planned Profile of Learning assessments, which will be state-developed classroom instruments, could be an interesting development in state assessments, though it is too early to know precisely what they will look like in practice, or whether they will be a successful innovation. The new legislation simply requires district testing and leaves the details to the SEA. The SEA should make certain that they involve performance or portfolio approaches and remain locally flexible. Equity issues will require very close monitoring as the new system develops, as will the curricular and instructional impacts and how the new assessments interact with the high-stakes BST. Extensive professional development and education of the public will be essential for success. If this component develops well, it might make an important contribution to national assessment practices, as well as to education in the state.

Standard 1: Assessment supports important student learning.

Minnesota is developing standards in several categories: Basic Requirements in reading, math and writing; High Standards in the Required Learning Profiles for arts, reading, writing, speaking, listening, math, social studies/history, science, problem solving, inquiry, and use of resources -- all to be at a complex or advanced level; and Standards of Distinction.

The state's Basic Standards Tests (BST), based on the Basic Requirements, are minimum competency tests to ensure that all students have basic literacy and numeracy skills. The BST is intended to guide curriculum and instruction. It includes criterion-referenced, multiple-choice tests in reading and math and an on-demand writing-to-a-prompt test, still in preparation. Students will be tested in grades 3, 5 and 8. The grade 8 math and reading tests can serve as the high school exit exam. An alternative assessment, however, can be used for this decision.

The Profile of Learning assessments, based on the High Standards for Required Learning Profiles, will be developed for use at the classroom and school levels. They are to be implemented by the 1999-2000 school year. The plan is to use a mixture of assessment methods. Results will be used locally for determining eligibility for graduation, with students required to receive a passing score in at least 24 categories, of which 18 are to be distributed among requirements in 10 subject areas. Within the 10 areas, students will have some flexibility to select a focus topic. The decision as to whether students have passed will be made locally. Based on suggestions made in the Profile of Learning standards documents, extended projects or portfolios may be included in the Profile assessments.

Teachers, administrators, SEA staff, outside experts, students, parents, community and education organizations were involved in designing the BST. The first three groups were involved in writing items, scoring rubrics and selecting examples of work. Scoring of the writing is done by outside experts.

Evaluation. Positively, the Profiles are to be based on standards, allow for substantial local flexibility, and include multiple methods. As they are under development, it remains to be seen how good they will be, including what proportion of items will be multiple-choice or short-answer rather than extended-response or performance items. It is not clear whether portfolios will be included. Negatively, the state requires a high school exit test, but, unusually and positively, it does allow for alternatives. It is not yet certain how the Profiles will be used for determining high school graduation. Currently, the testing burden is not heavy, but depending on how they are used, the Profiles could be rather overwhelming. The state should develop the Profiles assessments carefully, ensure that there is substantial local control over their use but monitor to prevent misuse, and ensure they include substantial amounts of extended response or performance tasks or a portfolio approach. The state needs also to ensure that the alternative to the BST high school test is readily available and useable for students, though it would be preferable to drop the requirement. It also needs to ensure that curriculum and instruction are not organized toward the BST in ways that undercut the Profiles. Indeed, as the Profiles are developed, the BST should be eliminated.

Standard 2: Assessments are fair.

Representatives of all stakeholder groups are involved in the bias review committee. Items are pre-tested and analyzed before and after testing. The committee can make recommendations about the assessment and has authority to delete or modify items. The new legislation now requires data to be reported by demographic categories.

Nearly 11 percent of the state's students have an IEP and about 2 percent are LEP. Under the new legislation, IEP students may be excluded if the student's plan states she or he is incapable of taking a statewide test and parents give approval. LEP students may be excluded if they have been in the U.S. for fewer than 12 months and special barriers exist, such as no written language or lack of a translator. Accommodations will have to be available. In the past, IEP and LEP students have not been included in regular reports, but now all those tested will be included in state reporting, though LEP students may be separately reported.

Evaluation. Bias review for the BST appears adequate. The Profiles could be quite positive in allowing real flexibility and variety based on high level standards. Preventing bias in the local assessments envisioned in the Profiles could be a challenge for the state. Inclusion also needs to be monitored carefully. While the new law is far more inclusive, it still does not meet the full requirements of the new IDEA for students with IEPs.

Standard 3: Professional development.

The state offers training in most areas of testing and assessment, but has no pre- or in-service requirements. The state has not surveyed assessment practices at the district, school or classroom level. Test information, including descriptions of methods, samples, scoring guides and examples, has been provided to teachers, administrators, students, parents, the community and policymakers.

Evaluation. Substantially more systematic professional development is almost certainly needed if the Profiles are to become a successful, decentralized, high quality, largely-performance assessment program. The kinds of local activities the Profiles seem to seek are likely to require not only professional development but creation of a culture of professional collaboration and substantial restructuring of schools, which in turn will need support and guidance from the state. Educators have played a major role in developing the BST and will need to be similarly involved in developing the Profiles.

Standard 4: Public education, reporting and parents' rights.

Results of the BST are released in about six months. The state has conducted a survey to determine what information parents or the public want and whether reports are understandable. Reports are only in English. Parents can review items after testing, and some items are publicly released.

Evaluation. Positively, the state is making an effort to find out what information the public wants and to allow some openness in testing. A major public education effort will be required to supplement the Profiles approach.

Standard 5: System review and improvement.

Since it is new, no technical studies or studies of the consequences of the BST have been done. The Profiles are not yet in place. They appear to be an effort to fundamentally alter how assessment is done within the state. Careful studies and use of the resulting information to refine the system will be necessary. This should include studying the interaction between the BST and the Profiles and whether the Profile assessments really do match the high levels of the Profile standards.

Minnesota responded to the full FairTest survey and sent various documents. This report also used CCSSO and AFT reports and a copy of the recent legislation. The state responded to a draft description prior to the new legislation.

MISSISSIPPI

Summary Evaluation.

Although Mississippi did not respond to the FairTest survey, data from other sources indicate that the state's assessment system needs a complete overhaul. The state relies too heavily on multiple-choice items, uses an NRT in grades 4-9, and has a high-stakes high school exit exam. Data are weak on the other standards, but it appears that while inclusion of IEP and LEP students is similar to many states, it needs improvement. Some professional development in performance assessment is provided, but it is hard to tell how much. Public reporting seems quite minimal. No data were available on system review.

Standard 1: Assessment supports important student learning.

Mississippi has standards in language arts, mathematics, science, social studies, the arts, health and physical education, and business and technology. The standards are embedded in curriculum frameworks.

The Mississippi assessment program currently includes norm-referenced testing (ITBS and TAP), the Functional Literacy Exam (FLE), and the Subject Area Testing Program (SATP). The commercial NRT is administered to students in grades 4-9 in language arts, math, and reading. The test includes some multiple-choice with student explanation and some short-answer items. The language arts and math frameworks are correlated to the norm-referenced test.

The criterion-referenced, multiple-choice Subject Area Testing Program (SATP) includes end-of-course tests in Algebra I (grade 8) and U.S. history (grade 11). These tests were developed from the frameworks. Algebra I includes some multiple-choice with student explanation and some short-answer items. Similar item types are under development for the US history exam.

The Functional Literacy Examination (FLE) is a criterion-referenced, multiple-choice test that includes writing samples and is used as a high school graduation requirement. It is first administered to students in grade 11, in math, reading, and written communication. Writing prompts are provided by the exam contractor. Students are given one hour to produce a writing sample with no revisions permitted.

The results of the FLE and the NRT are used for curriculum improvement, program evaluation and school accreditation. The NRT results are also used for student diagnosis and placement, and for possible sanctions such as warnings, probation or school take-overs.

The FLE may be replaced with a new set of exams. The state is piloting use of ACT Work Keys and ACT Occupation-Specific Assessments. An end-of-course biology exam was scheduled to be piloted in 1996-97.

Evaluation. Mississippi's use of NRTs, testing in too many grades, over-reliance on multiple-choice, a high school exit exam plus high stakes for schools makes this a program that needs a complete transformation. The use of constructed-response items is a start, but should be greatly expanded. The NRTs should be dropped or reduced to a minimal sample. The end-of-course exams should not be used to determine whether a student passes (if they are so used) and should be primarily constructed-response. The high school exit test should be dropped.

The writing to a prompt is too brief to be a good measure of writing capability. Data were not available as to how test results were used in school evaluations and sanctions, but test scores should not be the sole basis for decisions.

Standard 2: Assessments are fair.

Video materials are available to inform students about the tests. No data on bias review was available. The state has an exclusions and accommodations policy for IEP and LEP students, which is applied individually to determine participation in the various testing components. All students must pass the FLE to graduate. A fairly wide range of accommodations for IEP students is available on the FLE, but few on the NRT or SATP. Exemptions for LEP students on the NRT are determined locally. A modest set of accommodations are available on the FLE and the NRT. No alternate assessments are available.

Evaluation. Based on the available information, the state needs to strengthen both inclusion of IEP and LEP students, including alternate assessments, and to report data by sub-populations as well as the entire tested group. Heavy reliance on multiple-choice items and the high school exit exam also do not meet this standard.

Standard 3: Professional development.

Print and video materials are prepared for professional development. A trainer-of-trainers model was used to prepare teachers for administration of performance assessments.

Evaluation. The trainer-of-trainers model can work well, but no data were available on the extent of the program and little data on the content.

Standard 4: Public education, reporting and parents' rights.

Print materials are provided to parents, but there none is reported as prepared for the general public or policymakers. Reports do not provide disaggregated data by any demographic characteristics or for IEP or LEP students. IEP and LEP students who were tested were included in regular reports.

Evaluation. Reporting seems minimal. The state noted that few people seemed to know about the state's standards. Disaggregation of data are needed. No information was available on parents' rights.

Standard 5: System review and improvement.

No data were available.

Mississippi did not respond to the FairTest survey. This report relied on CCSSO/NCREL for 1995-96, CCSSO and AFT reports.

MISSOURI

Summary Evaluation.

Missouri is undergoing a major shift in its assessment program to one that will only need modest improvement. The shift from multiple-choice to a mixed-method assessment is positive, the ending of sampling is not. However, since most districts used state assessments to test more than the new assessments will require, the shift away from sampling may not result in an actual increase in testing for students. Stakes are moderate. Bias review and inclusion need strengthening. Professional development appears to be quite strong. Public education and reporting are currently adequate but will need strengthening as new assessments are implemented. The review process needs improvement.

Standard 1: Assessment supports important student learning.

Missouri has content standards in communication arts, math, science, social studies, arts, and health/physical education, as well as interdisciplinary process standards. Curriculum frameworks based on the standards that are to be used as models for LEAs also have been adopted. A new assessment system is being developed based on these standards.

Currently, the Missouri Mastery and Achievement Test (MMAT) is available for district use in grades 2-10 in the areas of language arts, math, science, and social studies. These are criterion-referenced and multiple-choice tests which are based on and aligned with the Missouri Core Competencies and Key Skills, a document that has been replaced by the new standards. For state-level data on the MMAT, the SEA collects a sample of 6000 students in grades 3, 6, 8 and 10, and a writing sample in grades 5, 8 and 11. Since the test is optional at the district level, the state selects individual schools to be assessed in the few districts which do not test.

A writing assessment also is voluntary for districts. For state data, a random sample of buildings statewide is drawn. It uses multiple-choice items and writing samples with SEA provided prompts in grades 5, 8, and 11. Students in the same grade level receive the same prompts and three class periods to produce a writing sample. Scoring is holistic and done by teachers in the state.

Plans are underway to phase out the MMAT in the next four years. New mandatory assessment components, the Missouri Assessment Program (MAP), to be based on the new standards and frameworks, will be implemented in math, reading, writing, science, social studies, theater, visual arts, dance, music, health education and physical education. The assessments are in various stages of development (from funded but not started, to piloted and being refined). One grade at each of three levels -- elementary, middle and high school -- will be assessed, with grades varying by subject so that no grade is tested in more than two subjects.

Each MAP component will employ a variety of exercise types. The pilot of the nearly-finalized grade 8 math test contained two performance events, 11 constructed-response items of varying length, and 31 multiple-choice items. About 55 minutes will be spent on each item type. It is expected that other subjects and grades will have similar proportions and times. (Testing per grade will therefore equal about 6 hours.) The scoring weights for the different parts have not been determined. Reporting will probably be of the population, not a sample.

The new writing assessment will be developed by a contractor, which will score it. Ten percent of the papers will be rescored by Missouri teachers.

MMAT design and item writing has involved most stakeholders except students, with education organizations as "observers." Development of MAP includes all stakeholder groups.

Results of assessments are used for student diagnosis or placement, but the state says they should only be one piece of information, not the sole basis for decision-making. They are also used for curriculum improvement and program evaluation. For schools, assessment results, reviewed on a longitudinal basis, may be used as part of determining accreditation, warnings or probation. Current plans do not call for accountability changes with the new assessments.

Evaluation. The current MMAT appears to be both burdensome (if districts test all the grades) and positive, in that Missouri is the only state to rely exclusively on sampling for state data. The MMAT, however, is all multiple-choice. The new assessments will be an improvement in methodology, and the burden is not heavy for any one grade, but unfortunately sampling has been dropped. The score weights for the different sections have not been determined; the weight of the multiple-choice should not reflect more than its share of testing time. Thus, the changes are generally positive. Stakeholder involvement is strong. The accountability requirements are moderate. Using assessment results longitudinally to measure improvement over time for accountability analysis is positive.

Standard 2: Assessments are fair.

A bias review committee exists and is made up of parents, business people, and community leaders of diverse backgrounds representing different areas of the state. Demographic representation is ensured for these groups. African Americans constitute a larger proportion of the bias committee than their proportion of the state population. The committee proofreads items and makes recommendations, but it has no authority to delete or modify items. Items are also analyzed before and after testing for bias. Reports do not include information by demographic categories.

Over 13 percent of the state's students have an IEP, while 9 percent of tested students do. Fewer than .5 percent are LEP. Both may be exempted from the assessments; IEP students may receive accommodations. IEP and LEP are excluded from regular reports. Accommodations on MAP are being discussed, and the emphasis will be on inclusion.

Evaluation. The review committee appears properly inclusive, but should have more authority. Reports should include demographic-based data. Much more will need to be done for accommodations or alternative assessments for IEP and LEP students. The development of multiple methods should assist equity.

Standard 3: Professional development.

Pre-service teachers are required to receive instruction in testing, observational techniques, and classroom assessment. For in-service, tied to MAP, the state uses a trainer-of-trainers model on developing, scoring and using performance assessments. Nearly half of the districts have been involved thus far. Administrators are surveyed on professional development needs in their schools and districts, including assessment. Annual teacher evaluations done at a local level include competence in assessment. SEA staff visit districts on a five-year cycle, at which time assessment practices are reviewed.

Evaluation. Pre-service requirements are positive, and the model for in-service is systematic and steadily covering the state. The surveys, evaluations, and site visits are also very positive, as is teacher involvement in assessment development.

Standard 4: Public education, reporting, and parents' rights.

On MMAT, old tests are available to be used as pre-tests. General printed materials on testing are available to students, teachers, parents and policymakers. Videos are available to all but students. A general video on performance assessment and MAP is available for schools, and a brochure is available for parents of students who will take the new tests. Information on assessment methods, samples of assessments, and scoring guides are provided to administrators, and to parents or community members, on request.

State public reports on the MMAT are based on the samples, with grade, school and district reports released. The state has not surveyed parents or the public regarding information they want or whether they understand the reports. Results are reported to students, parents, and schools in 4 to 8 weeks and to the public in 8 to 12 weeks, in English only. The SEA provides guidance on the use of results to all.

Parents may opt out of testing and (with safeguards) review items after administration. Parents or students can appeal a score or challenge items as flawed.

Evaluation. The state's public education efforts are good, but more will be needed as the state moves more towards performance assessment. Permitting parents and students to review and appeal test items is positive.

Standard 5: System review and improvement.

The MMAT is evaluated by the SEA. The MMAT is intended to guide curriculum and instruction, and the state has found that the key skills and core competencies on which the test is based are embedded in districts' curricula, but the test has not been evaluated for its impact. The MMAT has not been evaluated for consequences or for whether it assesses critical thinking, but it has been reviewed for developmental appropriateness. Technical studies of reliability and measurement error have been done on the tests. Review processes for the MAP have not been determined.

Evaluation. Review of the MMAT was not sufficient. The evaluation of MAP should be stronger and comprehensive.

Missouri responded to the full FairTest survey. This report also relied on CCSSO/NCREL, CCSSO and AFT reports. The state responded by telephone to a draft descriptive report.

MONTANA

Summary evaluation.

Montana has a bare-bones state assessment program that needs many major improvements. The state system relies entirely on multiple-choice, norm-referenced tests in three grades. This program should be replaced. Districts are allowed to choose from a list of approved NRTs. They are also required to develop additional assessments. This optional approach should be built upon to help districts implement primarily performance assessments in key subjects. Since the state relies on the NRTs for state-level data, one option would be to continue to do so, but on a sampling basis. A preferable alternative would be to lead a collaborative effort to develop a state rubric that districts may use to score performance assessments or portfolios. This rubric could be the basis for rescoring district samples for state-level information.

Little is done to ensure proper assessment of IEP or LEP students, nor does the state address bias. The state has no professional development program or requirements for teacher competence in assessment. Districts are also responsible for reporting to parents. The state is only now considering a review of the state's program. Since Montana's political culture strongly favors local control, the state has more limited options for promoting reform. In addition to helping districts implement new assessments, establishing requirements for incoming teachers and offering encouragement and guidance in other areas of equity, professional development, public education and reporting would be one route toward strengthening the state's assessment program.

Standard 1: Assessment supports important student learning.

Montana's Standards for Accreditation of Schools includes nine program areas in which districts develop their own standards and curricula. *Model Learner Goals* provide guidance in communications arts, math, science, social studies, fine arts, health enhancement, vocational/practical arts, library media and guidance, for elementary, intermediate and high school levels.

The Board of Public Education approves a list of five standardized, off-the-shelf, norm-referenced tests (CAT, CTBS, ITBS, MAT, SAT) from which districts must choose (called the Student Assessment Requirement). Students are tested in grades 4, 8, and 11 in the areas of language arts, reading, math, science and social studies. Scores are summarized at the state level by grade, test and subject. State rules prohibit the state from comparing districts or schools. Results are used for curriculum improvement and program evaluation by districts, which may use the tests for other purposes at their discretion.

Districts are required to develop assessment methods for each area of the curriculum, in addition to the state tests. The kinds and quality of assessments used is not studied.

Evaluation. The state should drop the NRT requirement, or use it on a sampling basis only, and help districts implement performance assessment programs. A state rubric could be developed collaboratively to enable state-level data to be obtained.

Standard 2: Assessments are fair.

Statewide, 8 percent of tested students have an IEP, and 4 percent are classified LEP. IEP students may be excluded from testing. IEP and LEP students who are tested are included in regular reports.

Evaluation. The state appears not to provide guidance to districts on bias and equity issues. The state does not list any accommodations for IEP or LEP students; if districts follow commercial test publisher procedures, a very limited range of accommodations may be allowed. This means that many IEP and LEP students are not adequately assessed or included in data. Requiring an NRT limits the opportunity for assessments to meet diverse learning styles and cultural backgrounds.

Standard 3: Professional development.

The state has no specific requirements for preservice or in-service teacher education in assessment. The state does not provide such education, nor survey teachers regarding needs for training in assessment. The state does not evaluate teacher competence in assessment.

Evaluation. As a state with significant local control, Montana appears to leave professional development to the districts. However, guidance in such matters, as well as establishing requirements for incoming teachers to be competent in classroom and performance assessments, would be reasonable steps for the state to take.

Standard 4: Public education, reporting, and parents' rights.

Any reporting of scores to parents or the public is done by districts. The state only releases a state-level summary, which contains no reporting by subpopulations.

Evaluation. The state should gather and report data by demographic groups and guide district reporting to the public and parents. If the state begins to use performance-based assessments, the state or districts should involve the public in developing the assessments and should provide public education about them.

Standard 5: System review and improvement.

The state does not have a regular review process, but currently the Board and Office of Public Instruction are developing a process to review the state-level assessment program. The state does not survey or evaluate district practices.

Evaluation. Hopefully, the review of the state system will lead to some positive changes. As the state leaves most assessment to districts, it would be reasonable to survey districts about their practices and provide support for improvement.

Montana responded to the full FairTest survey. This report also relied on CCSSO/NCREL, CCSSO and AFT reports. The state responded to a draft descriptive report.

NEBRASKA

Summary evaluation.

Nebraska has a minimal state program that needs many major improvements. The state system requires districts to administer an NRT in at least three grades. This program should be replaced. Districts also are required to develop additional, criterion-referenced assessments. This approach should be built upon to help all districts implement primarily performance assessments in key subjects, as some are now doing. If a state exam is developed, as is required by law, it should be implemented in coordination with local assessments. It should be a high-quality assessment that fully reflects the new state standards and is not overly reliant on multiple-choice items. The state should guide districts to ensure proper bias reduction is used. The state should provide substantially more support for professional development in assessment. It also should evaluate the impact of its mandates on curriculum and instruction and the actual local assessment practices, and it should provide support at the local level, as needed, for developing performance assessments.

Standard 1: Assessment supports important student learning.

In 1992, the legislature required the Nebraska Schools Accountability Commission to develop curriculum frameworks, standards, assessments and a state accountability system within four years. As funding has not been adequate, this is still in development under auspices of the SEA. The State Board of Education has approved drafts of frameworks in science, math, social studies, reading and writing, and they are scheduled for public discussion. The board plans to affirm final versions by late summer, 1997.

Nebraska does not have a statewide assessment system. "The current accountability system requires reporting at the lowest level of jurisdiction to parents and patrons." The state requires all districts to use a NRT at least once in grades 4-6, 7-9, and 10-12. No subject areas are required, but most districts use a full battery. Some include writing to a prompt.

Districts also must begin to gather criterion-referenced assessment data beginning in grade 5 in reading, writing, and math, and continue in subsequent grades and other subjects according to a district plan. These assessments are based on local checklists or benchmarks of progress, student portfolios or other criterion-referenced measures.

The statewide curriculum frameworks are intended to guide development of local frameworks and assessments. Legislation has been introduced to develop statewide tests in grades 3, 7 and 10 beginning in 2000, thereby implementing the earlier law.

Evaluation. Negatively, the state requires use of an NRT. More positively, it requires districts to gather criterion-referenced assessment data. This requirement could allow multiple methods of assessment and substantial performance assessment, but FairTest has no information on the extent to which districts do that. The intent to have state frameworks guide local frameworks and assessments is a positive approach, but needs to be monitored and supported. If a statewide test is approved and added to other mandated testing, it will produce an unreasonable testing burden -- yet another reason to drop the NRT.

Standard 2: Assessments are fair.

Any bias review procedures for local assessments are done locally. Bias review on the NRT is done by the contractor. Locally developed reports must include demographic data.

Decisions on assessment of IEP or LEP students are made locally, and districts vary in their practices. There are no state policies governing such decisions, but state committees are currently working on state guidelines, which are expected to be promulgated by fall 1997.

Evaluation. The state should monitor local practices for bias. Incoming state guidelines on assessing IEP and LEP students should be helpful. Requiring an NRT limits the opportunity for assessments to meet diverse learning styles and cultural backgrounds. Positively, the state requires extensive reporting by districts, including demographic information.

Standard 3: Professional development.

The state has no requirements for knowledge in assessment for incoming teachers. Service units at the regional level provide some assessment support under the school improvement portion of accreditation. The state has not evaluated assessment practices at the classroom, school, or district levels.

Evaluation. Professional development in assessment appears inadequate, as there is neither state-provided training nor support for systematic district-provided training, and the state has no requirement for incoming teachers.

Standard 4: Public education, reporting and parents' rights.

All accredited schools or systems must include in their annual reports to district residents data on student achievement, demographic information, school climate, graduation follow-up studies, and financial information. The state monitors compliance with this requirement through audits and does not collect district achievement data.

Evaluation. Positively, the requirements for reporting are fairly extensive. The clear state intent is to have substantial local control over and local participation in assessment, and to provide extensive local reports. Limited information is available about the actual quality of the reports, the extent to which local educators are involved in developing or scoring the local assessments, and how much the public is educated about assessment practices.

Standard 5: System review and improvement.

Since there is no state system, the state does no system evaluation. While it monitors compliance with state policies through audits of LEAs, it does not evaluate local practices.

Evaluation. Evaluation of local assessment practices should be carried out beyond compliance audits. Studies of the impact of local assessments on practice and learning should be included.

Nebraska responded over the telephone to some questions from the short FairTest survey. This report also relied on CCSSO/NCREL, CCSSO and AFT reports. The state responded to a descriptive draft.

NEVADA

Summary evaluation.

Nevada's assessment program needs many major improvements, primarily shifting from multiple-choice to predominantly performance assessments and eliminating the high school exit exam requirement. The state has a rather light assessment burden, testing only a few subjects in three grades. Bias review and inclusion of IEP/LEP students should be strengthened. Professional development needs substantial strengthening. Public reporting appears adequate, but a survey should be done to confirm this. The review process needs improvement. In particular, the state should evaluate the impact of assessment on curriculum, instruction and graduation, and should study the ability of the assessments to measure critical thinking and cognitively complex work.

Standard 1: Assessment supports important student learning.

Nevada has state standards in reading, math, writing, science and social studies. The Nevada state assessment program includes the state-developed Nevada High School Proficiency Examination (NHSPE), first administered in grade 11, in which math and reading are tested with norm-referenced, multiple-choice exams. Writing is assessed with responses to SEA-provided prompts (one hour to answer two prompts). A multiple-choice NRT, the CTBS/5, is used to assess students in grades 4 and 8 in the areas of math and reading and grade 4 in the area of language arts. A grade 8 writing assessment uses an SEA prompt to stimulate production of a writing sample over two 35-minute periods on consecutive days.

In 1997-98, the state will introduce criterion-referenced, multiple-choice tests at grade 11 in math and reading to be used as the graduation test. Students will be able to retake the exams through grade 12. These will replace the NRTs which have been in place since 1990.

The purchased NRTs for grades 4 and 8 are aligned with the state's standards according to studies by the publisher. The writing assessment is aligned with content standards through the scoring rubric. The graduation exam contains items written to specific objectives in a state course of study. The SEA recognizes there are areas of the standards within tested subjects that cannot be tested through multiple-choice items and the writing sample.

Results of assessments are used for student diagnosis or placement, curriculum improvement and school performance reporting. Results are not intended to guide curriculum and instruction. The High School Proficiency Examination Program results are also used as an exit requirement. Students who do not pass one or more of the graduation tests can take those parts up to four times in grade 12. After grade 12 they can continue to take the necessary tests on presentation of evidence of additional remedial study. There are no high-stakes consequences associated with the results of other assessments. Districts include test results in district accountability reports that they are required to produce and distribute.

Evaluation. Shifting from NRT to CRT testing at grade 11 is positive, but reflects minimal progress. The state relies entirely on multiple-choice, except writing to a prompt, and needs to shift to a primarily performance assessment system. The graduation requirement should be eliminated. The most positive factor may be that the state tests very little.

Standard 2: Assessments are fair.

The state provides preparatory materials on the writing assessments to students. The Terra Nova is reviewed for bias by the publisher. Although the state does not have a standing bias review committee, ad hoc advisory committees are formed by the SEA. The SEA reports that assessment development has attempted to take account of different cultural backgrounds of the students, but not of different learning styles. Disaggregated data are now released by gender, SES and ethnicity.

IEP students can be excluded at grades 4 and 8 if their IEPs require exclusion, and LEP students can be excluded if English proficiency as measured by the Language Assessment Scales test is too low. Fairly extensive accommodations are available on the writing tests and the high school exit exam, including allowing students with LEP up to twice the regular time to take an exam. Limited accommodations are available on the grade 4 and 8 NRTs as determined by the publisher. Both IEP and LEP students must pass the HSPE to obtain a Standard Diploma. IEP students who do not pass can earn an Adjusted Diploma, while LEP students who do not pass can earn a Certificate of Attendance.

Evaluation. Bias review appears to be insufficient. Reporting disaggregated data is a positive development. The state will need to make major changes to meet the new federal IDEA requirements for students with IEPs. Reliance on multiple-choice items and the graduation requirement do not meet the fairness standard.

Standard 3: Professional development.

Nevada has no requirements for preservice training in assessment for teachers. It does not evaluate teacher competence in assessment, nor does it survey educators regarding their professional development needs in assessment. The state provides math and reading in-service programs that cover assessment, including scoring writing. Teachers score the writing assessments using rubrics developed by a state advisory committee.

Evaluation. The state needs to address professional development in assessment, which it largely has not done. Teacher involvement in scoring is positive.

Standard 4: Public education, reporting and parents' rights.

It is the district's responsibility to determine what information to report to parents and the public, with a number of elements, including test scores, required by law. The state has surveyed the districts in deciding what CTBS/5 reports to use. Print materials are provided to parents. As this is the first year it has administered the CTBS/5, the state has not surveyed the public to determine whether the reports are understandable. They are released in English only.

Parents can exempt their children at any tested grade, with the understanding that the student will not earn a standard diploma if he or she does not take and pass the graduation exam. Students may appeal scores and challenge items as flawed (both are rare occurrences). Parents cannot review assessment items after the exam is completed, except for the writing assessments in grades 8 and 11/12. In that case, parents can review their child's response to prompts under the guidance of a teacher, who can explain the scoring process and identify the student's weaknesses.

Evaluation. Reporting is probably adequate for the limited state program, but a survey would be useful to confirm this. Parents should be able to review tests as this is done with commercial exams in some other states. The rights to appeal and challenge are positive.

Standard 5: System review and improvement.

The SEA reviews the state assessment system annually. Alignment between state standards and the exams has been evaluated by the SEA and district representatives. The state has technical studies for the writing assessment, which shows that scoring is quite reliable. The SEA does not study the impact of assessment on curriculum, instruction or high school graduation rates. The assessment has not been evaluated for how well it measures cognitive complexity or critical thinking. The SEA has evaluated assessment practices at the district level, but not at the school or classroom levels.

Evaluation. Stronger and more extensive reviews are needed, focusing on three things: the impact of testing on curriculum and instruction to see if, despite intentions, the tests do affect them; the impact on high school graduation; and the capacity of the assessments to measure critical thinking and cognitively complex work in the tested subject areas.

Nevada responded to the short form of the FairTest survey. This report also used CCSSO/NCREL, CCSSO and AFT reports. The state replied to a draft description.

NEW HAMPSHIRE

Summary evaluation.

New Hampshire's state assessment program needs modest improvement, primarily shifting the balance from majority multiple-choice to predominantly constructed-response and performance assessment. The light testing burden and relatively low stakes are positive. The state does well on all other principles, with either minor improvements or expansion of existing efforts advised.

Standard 1: Assessment supports important student learning.

New Hampshire has curriculum and assessment frameworks in math, English language arts, science and social studies, developed by educators, business people, government officials, community representatives and parents.

New Hampshire's assessment program, the NH Educational Improvement and Assessment Program (NHEIAP), consists of CRTs based on the frameworks that use multiple-choice and short-answer, open-ended items and writing samples. Students in grade 3 are tested in the areas of English language arts (ELA) and math. Students in grades 6 and 10 are tested in the areas of ELA, math, science and social studies. In ELA, a writing sample comprises 30 percent of the score, multiple-choice 45 percent, and constructed-response 25 percent. In other subjects, multiple-choice is 60 percent and constructed-response 40 percent. All students are tested and multiple forms are used. Matrix sampling is used for some items on the exams. Areas of standards that cannot be tested with paper and pencil are not assessed.

Writing assessments use SEA provided prompts. All students have a minimum of 70 minutes to complete a writing sample; students who are still working are given an unlimited amount of additional time. Revisions are permitted. Scoring is done by the test contractor using a rubric developed by the contractor with SEA staff and a content committee.

All stakeholders, except students and advocacy groups, are involved in committees charged with developing test content or reporting results. Teachers comprise the majority, accounting for 150 out of 200 members.

Results of assessments are used for curriculum improvement and school performance reporting with no high-stakes consequences for either schools or students.

Evaluation. Though the state rates fairly high in comparison with most states, New Hampshire's program still relies too heavily on multiple-choice. The relative balance of methods should at least be reversed. The light testing burden and relatively low stakes are positive, as is permitting extended time for the writing assessment.

Standard 2: Assessments are fair.

Bias review is carried out by the content committee, which can modify or delete items. Statistical analyses are conducted before and after administration. Use of different formats and provision of accommodations are used to respond to different learning styles. Content committees and the contractor review the assessment for developmental appropriateness. The SEA provides sample tests for grade 3, and curriculum frameworks and released items for other grades, to provide test familiarity for students.

Fourteen percent of all students and 11 percent of students tested have an IEP. Fewer than 1 percent of students are categorized as LEP. Extensive accommodations are available, but these do not include assessments in languages other than English. Limited numbers of IEP and LEP students are excluded from the assessments. Results are included in regular reports. State, district and school data are produced by gender, enrollment in Title I, LEP or IEP.

Evaluation. The bias reduction efforts are solid, as is the use of multiple methods in the assessments and the reviews for developmental appropriateness. Accommodations appear solid and a higher proportion of students with IEPs are assessed than is the case in most states. Reporting is good.

Standard 3: Professional development.

NH requires preservice knowledge by teachers and administrators about standardized testing, classroom assessments, and the use of test results. Further professional development is offered, but not required, by the state. Samples of items and scoring guides are used for explaining the state assessments to educators. Training is provided in administering assessment accommodations, understanding reports and using test results. Seventy-five percent of districts have requested training. The state does not evaluate teacher competence in assessment or survey teachers to determine if their professional development needs are being met.

Evaluation. Professional development is decent but can be strengthened with continued in-service education in classroom assessment practices and by surveying for teacher competence and needs. Scoring of writing samples and any extended-response items should be done by teachers.

Standard 4: Public education, reporting and parents' rights.

Samples of items and scoring guides are used for explaining the assessments to parents, policymakers and the community. Published reports are released five months after testing in English only. They explain the four proficiency levels (or, for writing, "commendations" or "needs" for 6 scoring areas) and report the percentage of students attaining each level. Results are reported by the state, district and school. A brief survey on instructionally-related topics is also administered to students and the results are reported at the school, district and state levels. The SEA has surveyed parents and the public regarding information they want and whether the reports are understandable. Parents and students over age 18 can review items once results are released.

Evaluation. The state appears to do a solid job of educating and reporting. There is some public involvement in developing the assessments, and the right to review is reasonable. Use of the various surveys is positive, as is reporting by various demographic categories.

Standard 5: System review and improvement.

Assessment review is ongoing, involving the SEA and independent evaluators. Evaluation includes the review of alignment with standards and the impact of the assessment on curriculum, instruction and education improvement. Technical studies, which include data on item-level difficulty, are provided by the contractor. Students receive a questionnaire which includes questions on learning and instructional techniques. About 30 percent of test items are replaced each year.

Evaluation. The evaluation process appears solid. Surveys of district assessment practices would be a useful addition to the review process.

New Hampshire responded to the full FairTest survey. This report also used CCSSO/NCREL, CCSSO and AFT reports. The state responded to a draft descriptive report.

NEW JERSEY

Summary evaluation.

New Jersey's assessment program needs many major improvements, including eliminating the high school exit requirement, shifting the emphasis from multiple-choice to constructed-response items, bringing the assessments in line with the standards, and substantially altering the state mandate for district testing. Because of the district mandate, the test burden is heavy. Bias review is solid, but inclusion should be expanded. Professional development needs substantial expansion and strengthening. Public education is currently adequate but will need to increase as the assessment program develops. Review needs some strengthening, including reviewing district assessments since the state mandates district level testing.

Standard 1: Assessment supports important student learning.

New Jersey has recently adopted content standards in language arts, math, science, social studies, world languages, arts, health/physical education, and cross content workplace readiness standards. A variety of groups, including educators, parents and business, were represented on the drafting and reviewing committees. Curriculum frameworks are being developed. Assessments will be aligned to standards.

The New Jersey assessment program tests math, reading and writing through the Grade 8 Early Warning Test (EWT) and the Grade 11 High School Proficiency Test (HSPT), which is a graduation test. Both tests are criterion-referenced and use multiple methods, including multiple-choice, short- and extended-response and open-ended items. All students in designated grades are tested and all students see the same items. Consultants, a commercial firm and committees that included parents, business reps, teachers and administrators were involved in developing the tests and writing items and scoring guides.

On the HSPT, each of four reading test parts has one open-ended question, while the math exams include some grid-in answers and some extended-response items that require students to construct a response and explain it. Open response counts for about 60 percent of the score in writing and about 25 percent in reading and math.

The writing assessments use multiple-choice items and responses to SEA prompts. Students are asked to produce a writing sample in 60 minutes for the HSPT 11 and in 40 minutes for the EWT. Scoring is done by a commercial testing company.

The SEA plans to administer a test to all fourth graders in 1998, beginning with math, science and language arts. Assessments in science and social studies will be developed for all three grade levels (4, 8 and 11), and students will have to pass these assessments to earn a diploma.

The state currently requires districts to administer a standardized achievement test from a state-approved list to all students in grades 3-7, 9 and 10. Once the grade 4 tests are in place, the state plans to modify this mandate and instead require districts to assess students annually using an instrument of the district's choice. No guidelines for the district-selected assessments have been established.

Results of the state assessments are used for student diagnosis, placement and remediation, curriculum improvement and program evaluation. Students may take the HSPT a maximum of four times in order to reach a passing score. For schools, test results are used for school performance reporting (school report card) and school accreditation. Consequences for schools may include probation, funding loss, accreditation loss and takeover, but test scores are not the sole criterion. For a district to be certified, 85 percent of eleventh graders must pass the HSPT and 75 percent must pass the EWT. A new law establishes rewards for schools if 90 percent of its students attain the test standards or the school makes unusual progress in raising scores.

Evaluation. New Jersey's assessment program has major problems, some of which may be adequately addressed with the planned changes. The graduation test requirement should be dropped. The balance of items should shift strongly toward open-response. The tests need to be brought into alignment with the standards, and the tests need to adequately assess the standards. The district testing mandate is also a major problem. It is likely to impose far too high a testing burden, even if the requirement is somewhat more flexible. Instead, the state should help districts develop classroom-based assessments, from which sampling can be done, and leave the large-scale tests to the state.

Standard 2: Assessments are fair.

A bias review committee (sensitivity committee), with demographic variety and including members of community groups, reviews tests for language, stereotypes, confusing context, socioeconomic/experiential background bias and gender bias. The committee approves tests and has the power to eliminate items, with input from the content committee. Items are analyzed pre- and post-testing, including statistical review. Equity is evaluated for each test administration. Disaggregated data are reported by SES, but not by race or gender.

Nine percent of tested students have an IEP and 3 percent are LEP. LEP and IEP students may be excluded; accommodations are available for IEP. The SEA says the grade 4 exam "is being built to accommodate special and disadvantaged students." If students with LEP do not pass the HSPT, they may take a special review assessment (SRA) in one of 10 languages, which involves performance and portfolio assessment. If they pass, they then must demonstrate some capacity in English language by obtaining a score of 133 on the Maculaitis test (which is also used for entry and exit from ESL programs). IEP students may also utilize SRA. IEP and LEP students are excluded from regular reports, and separate reports are issued.

Evaluation. Bias review is solid. Reporting should provide further disaggregation, but IEP and LEP should be included in regular reports. Requiring the high school exit test does not meet this principle, but, positively, alternatives do exist for IEP and LEP students. We do not know how many students this actually helps to graduate. When tests are revised to meet the new standards, they should, like grade 4, be built to accommodate IEP and LEP students. The use of open-ended items helps meet the need for assessments that respond to diverse learning styles and cultures, but more open-ended items are needed.

Standard 3: Professional development.

The state requires no professional knowledge of assessment beyond requirements for initial certification. It has not evaluated teacher competence in assessment nor surveyed educators for their professional development needs. Teachers receive training in understanding and using test results. The SEA plans to start professional development when frameworks and revised assessments are ready.

Evaluation. Professional development is currently inadequate for both pre- and in-service teachers. Surveys should be done. Professional development should extend beyond the state's revised assessments and include classroom assessment. Educator participation in test development is positive, but teachers should be involved in scoring writing and extended-response tasks.

Standard 4: Public education, reporting and parents' rights.

Students and parents receive an informational pamphlet about the SEA tests. Scores can be appealed, but items cannot be challenged (secure test), and only essay prompts can be viewed after the test. The state has not surveyed parents or the community about assessment issues.

Results are reported, in English only, within 2-3 months to students, parents, and schools and through a summary report to the public approximately every November. Guidance on the use of the results is provided to all.

Evaluation. Public education is not sufficient, and parental rights should be expanded. Reports possibly should be in languages other than English. With new assessments, public education and reporting should be strengthened.

Standard 5: System review and improvement.

The state tests are reviewed regularly by the SEA. Consequences of the previous high school exit test were reviewed. Of those who did not pass, "the vast majority did not graduate because of attendance and/or lack of completion of course requirements." Student scores have gone up over time. The state recognizes that the test will affect curriculum and instruction. The SEA does not survey districts or schools about their assessment practices.

Evaluation. Reviews should be strengthened regarding the impact on curriculum and instruction -- rising scores may be real learning gains or artifacts of test familiarity. As the requirements for the districts change, the state should monitor district assessments to ensure they support important learning, are not too burdensome and meet principles for fairness.

New Jersey responded to the full FairTest survey. This report also used CCSSO/NCREL, CCSSO and AFT reports. The state responded to a descriptive draft and send test specification reports.

NEW MEXICO

Summary evaluation.

New Mexico's assessment program needs a complete overhaul. The key positive steps the state could take are to drop the NRT and the high school exit exam and replace them with a standards-referenced, largely constructed-response state assessment program in three grades. Planned new assessments in grades 4, 6 and 8 could be a step in this direction, but will create an even heavier test burden. It also appears they will remain predominantly multiple-choice and short-answer. Positive attributes, such as teacher involvement in assessment development, assessments in multiple languages, and extensive evaluation should be continued and expanded. Increased professional development is also a must, as is expanded public education. The review process is positive.

Standard 1: Assessment supports important student learning.

New state goals, content and performance standards, and assessment and curriculum frameworks, in language arts, math, languages, arts, science, social studies, math and other subjects, will replace current frameworks. In the next two years, Assessment Blueprint 2000 will start aligning assessment with standards and benchmarks as they are established. Meanwhile, the SEA reports that assessments are aligned to the old frameworks.

New Mexico currently administers four assessments, and is developing new ones.

A norm-referenced standardized test, the ITBS (form K), for grades 3, 5, and 8 in the areas of math, reading, spelling and vocabulary is used for accountability and school reporting. (This will be changed to grades 4, 6 and 8 in 1997-98.) Test information is provided to schools by the contractor. The test is intended to guide curriculum and instruction; alignment has been analyzed by publisher. The test is not recommended for making decisions about students, but solely for program evaluation. No school level consequences are attached.

The criterion-referenced High School Competency Examination (HSCE) is given to tenth graders in language arts, math, reading, science, social studies, and writing. Passage is required for graduation; alternatives are available to IEP and LEP students. Students have the option to retake the HSCE at grade 11 or 12 a maximum of four times prior to exit from high school, and five years after high school. The subject area tests use multiple-choice (90 percent), and short and longer open-ended items. Writing is to an SEA prompt, with two hours to respond. While the assessment is aligned with previous state standards, the state recognizes it is inadequate for measuring some aspects of the new standards, and it will be revised.

Writing is assessed through portfolios in grades 4, 6 and 8 (optional for eighth graders; about 1/2 are tested), as well as by the HSCE writing test. The writing assessment is used for diagnosis, improvement of curriculum and instruction, program evaluation, and school reporting. Three prompts (narrative, persuasive and expository) are delivered to teachers in the fall. Students work toward them over the next four months. Teachers and students jointly select the one best piece, which is sent to a private company for scoring using a rubric developed by teachers, the SEA and a contractor. No individual or school stakes are attached to this assessment.

A mandatory Reading Assessment for Grades 1 & 2 is designed by districts, typically involving teachers and administrators, and reported to the SEA. It is intended to be aligned to state standards. It may be used for student diagnosis, placement or grade promotion, and it is used for program evaluation and school performance reporting. Responsibility for grade level appropriateness rests with the district.

The state has released Requests for Proposals for new assessments in grades 4, 6 and 8. These require all items to be aligned with a content standard. Educators in the state will check the alignment. The assessments will include multiple-choice, which will yield some normative comparisons, and constructed-response, criterion-referenced items. As resources become available, the SEA hopes to include extended tasks.

Evaluation. Reliance on a basic skills, multiple-choice NRT as the major basis for accountability and for guiding curriculum and instruction is a major negative. The test may be aligned with the current frameworks, but the frameworks would only be aligned with the test if they were extremely weak and low-level. If the new standards are an improvement, then incompatibility with the NRT will increase. Also negative is the use of a high school exit exam. The use of portfolios for writing assessment is a positive step beyond what most states do, though the concept of the portfolio, with mandated prompts, is limited. The reading requirement in grades 1 and 2 provides an opportunity for using high-quality classroom-based assessments, but what is actually done varies across districts. The state should evaluate and help improve district assessments as needed. The new assessment plan seems a marked improvement over the reliance on NRTs, but it does not appear that the NRTs will be eliminated. The result will be a heavy test burden as well as the continuation of tests that should be dropped. In sum, the state should use the new standards as the impetus to drop the NRT and implement new assessments, eliminate the high school exit test, allow more flexibility in the writing portfolios and extend them to the high school level, and ensure a classroom-based approach to the grades 1 & 2 reading assessment.

Standard 2: Assessments are fair.

On the state-made assessments, bias review committees are empowered to eliminate flawed items, statistical analysis is employed, and item writing committees deliberately involve educators from diverse cultures. For the grade 1 & 2 reading assessment, LEA's address the issue of bias. The publisher does bias analysis for the NRT. Reporting does not include result breakouts by demographic categories. Print information about assessments is provided to all stakeholders, reaching students through the schools.

Modifications, waivers or exemptions for all testing programs are allowed for IEP and LEP students, as needed, with different assessments using different policies or procedures. Twenty-three percent of the state's students are LEP; data about the proportion of students who have an IEP or percentages tested who have IEP and LEP are not available. The HSCE is available in Spanish, and other languages may be used in the writing portion.

Evaluation. Bias review efforts are positive, as are uses of accommodations, including allowing assessment in languages other than English. Over-reliance on multiple-choice, the NRT, and the HSCE are negative, as is the failure to report data by demographic categories.

Standard 3: Professional development.

Print materials for professional development are available to teachers and administrators. Information about assessment methods and scoring and samples of previously completed assessments are shared with teachers and administrators. Professional development on state assessment programs and the use of test results is offered to teachers, school administrators and other school personnel, but not required of incoming teachers. The state also conducts workshops on writing assessment. The SEA evaluates teacher competence in assessment, but it does not survey for professional development needs. Teachers, administrators, SEA staff and outside experts are involved in developing and scoring the state assessments.

Evaluation. Positively, the state encourages substantial teacher involvement in developing the HSCE, the writing portfolios and possibly the grades 1 & 2 reading assessments. Teachers score the writing samples for the HSCE, but not the writing portfolios, which they also should do. Professional development that can be useful for classroom assessment needs to be expanded and strengthened.

Standard 4: Public education, reporting and parents' rights.

Results on all state assessments are reported to students, parents, and schools in two months and to the public in six months. Results on the HSCE are available in both English and Spanish; the others only in English. The SEA provides guidance on the use of results to school administrators, psychologists/counselors and district administrators. For the Reading Assessment in Grades 1 & 2, results are reported by individual districts only to the state at the end of the school year. The state has not surveyed parents or the community about what assessment information they want or if they understand the information provided. Parents have limited opportunities to review tests, and students have only limited opportunities to appeal scores; on the HSCE, they can appeal only on the writing portion.

Evaluation. The HSCE report is available in Spanish, though others also should be, and reports possibly should be available in major American Indian languages. Public information should be expanded, particularly if new assessments are implemented. Parental review of exams and the rights of students to appeal possible flawed items should be expanded.

Standard 5: System review and improvement.

The SEA has a Statewide Evaluation Advisory Committee, comprised of teachers, administrators and SEA staff, which evaluates state assessment programs. Annually, it considers the impact of assessment on curriculum, assessment and high school graduation rates, but it has not conducted formal impact or consequential validity studies. For example, anecdotal evidence exists on the positive impact of the writing assessment on curriculum and instruction. The state has not surveyed district, school or classroom assessment practices, including the local reading assessments.

Evaluation. Positively, evaluation appears extensive at the state level, with significant involvement from educators. However, the evaluations do not include formal studies of the impact of assessment on curriculum and instruction. The quality of local evaluations of the grade 1 and 2 writing assessments is not reviewed by the state, nor has the state evaluated local assessment practices, a task it should consider doing.

New Mexico responded to the full FairTest survey. This report also used CCSSO/NCREL, CCSSO, AFT and GWU reports. The state responded to a descriptive draft.

NEW YORK

Summary evaluation.

The current system needs many major improvements. The state is overhauling its assessment system based on new standards. This will solve some, but not all, of the problems. The new assessments will utilize multiple methods, but the balance of methods is not yet determined. The exams should be predominantly performance and constructed-response. The Regents exams should determine only part of course grades, not whether a students passes the course, as is planned. The effect of requiring passage of these tests for passing courses means that, added together, the exams determine high school graduation. The testing burden in high school will remain very heavy, as will testing in the two other grades selected (4 and 8). Expanding assessment in other languages should proceed. More extensive professional development, in line with the new state assessments, the new local assessments and classroom assessment needs, should be supported. Extensive public education will be required. The test burden, the impact of new assessments on curriculum, instruction and high school graduation, as well as the match between assessments and standards, all will need to be carefully evaluated.

Standard 1: Assessment supports important student learning.

New York State is developing performance standards in all basic areas (some are completed), and the alignment of assessments to the standards is in progress. The state assessment program is undergoing extensive changes.

As of 1996-97, the state assessment program consists of numerous components:

-- Pupil Evaluation Program (PEP) Tests, multiple-choice CRTs that assess students in grades 3 and 6 in reading and math and grade 5 in writing. These are used for student diagnosis and remediation for low scoring students, among other things (see below).

-- Program Evaluation Tests (PET), multiple-choice CRTs with performance tasks (science) and essays (social studies) to assess students in grades 6 and 8.

-- Preliminary Competency Tests (PCT), multiple-choice reading CRTs with writing samples and some open-ended items, used to assess students in grades 8 or 9 who score below the median on the last PEP test in reading or writing, for diagnosis and remediation prior to RCTs.

-- Regents Competency Tests (RCT), primarily multiple-choice CRTs, with writing samples and some constructed-response items, to assess students in grade 9 in science and math, grades 10 and 11 in social studies, and grade 11 in reading and writing. Passing the tests is required for graduation. Obtaining a sufficiently high score on a specified Regents Exam, SAT or ACT test can be substituted for an RCT. Recent transfer students can receive exemptions under specified conditions.

-- Regents Examination Programs, 16 CRTs from majority to all multiple-choice, with writing samples and some performance or open-ended items, that assess students at the high school level in English, math, science, social studies and foreign languages. Used to award Regents' diplomas and honors diplomas.

-- Occupational Education Proficiency Examinations (OEPE), multiple-choice CRTs, some with short-answer responses, administered in grades 9 - 12. Passing the Introductory test and one other is required for a Regents' endorsed occupational diploma.

All these state tests are used for improvement of curriculum and instruction, program evaluation, public reporting, possible intervention into schools, probation or watch lists (PEP, PCT and RCT only) and staff accountability in the form of teacher awards or recognition and teacher evaluation or certification.

Second Language Proficiency Examinations in 5 languages are offered to students in grades 7, 8 and 9 who wish to earn one unit of high school credit for the study of a second language in elementary or middle school.

Reading tests use the Degrees of Reading Power (DRP) cloze method. All other tests are developed in-house with participation by selected teachers who draft items and review overall test construction. Writing assessments use SEA provided prompts and allow one hour plus to produce each piece. Scoring uses a rubric developed by the SEA with teachers, university, staff and researchers, and is done by teachers.

The Board of Regents has approved the basics of a new assessment system to be based on state standards. The current plan is to have criterion-referenced exams using multiple methods (multiple-choice, short-answer and extended-response items, and some performance tasks) in English language arts, math, social studies, science, and languages other than English (all required for graduation). The Regents will require local assessment (using state-developed tasks or guidelines, but not reporting to the state) in arts, health and physical education, career development and technology education. Testing on state exams will be at grades 4 and 8 and at various grades at the high school level. Other grades may be tested, still at the three levels (elementary, middle and high school), for local assessments. Subject-specific Regents exams for high school will be continued, but many will be revised to include more constructed-response items. Career-major exams will replace the Occupational Education Proficiency Examinations.

Passage of exams at the high school level will continue to be required for graduation. For grades 1, 2 and 3, local assessments in literacy to identify students in need of intervention will be required.

Pilot testing of the new assessments will begin as early as spring 1997; initial administration will be staggered from 1999 to 2001. LEAs may seek approval for alternative exams that are equally rigorous as the new state tests, valid and free from bias. The Regents exams will be developed as 2-to-3-hour tests, but will be administered in two blocks of 3 hours to allow more time for those who need it. Many details about the tests remain to be worked out as the exams are developed.

In designing a new system, the state is attempting to ensure the assessments adequately reflect and assess the standards. The state seeks a balance between extended and performance tasks and score reliability. A technical review group has been providing advice.

Evaluation. The current state assessment program is not based on standards, relies far too heavily on multiple-choice tests, imposes a somewhat excessive testing burden, and requires a high school exit exam. Positively, it does not use any NRT.

The new program will be an improvement, but not entirely so. The new assessments will not be so heavily multiple-choice, but the proportions are not yet certain, and changes in the high school exams may not be as extensive as in the tests for grades 4 and 8. High stakes will remain with tests used as a graduation requirement. Instead of having the test as a sole hurdle for passing a course (which in the past has had the effect of narrowing curriculum and instruction) end-of-course exams should count for only a part of a student's grade. Allowing alternative assessments at the local level may enable bottom-up creativity and a lessened emphasis on one-time tests.

Standard 2: Assessments are fair.

The SEA employs a bias review committee which has the power to discard items. Bias review is also part of the charge of content selection committees. The state does not report test results by demographic categories.

Thirteen percent of students tested have an IEP and 3 percent are LEP. Extensive accommodations are provided for both IEP and LEP students on all tests. On PEP and PET, IEP students can be excluded based on their plan. LEP students who have had less than two full years of English instruction can be exempted if the test is not available in their native language (math only). Partial exemptions are allowed for LEP students who entered an English-based program after enrollment in high school. Otherwise, IEP and LEP students must pass to receive a diploma. IEP and LEP students are included in regular reports, except IEP students on the OEPE. The Commissioner has proposed creating Regents exams in multiple languages for subjects other than English.

Evaluation. Bias review efforts appear reasonable, as do efforts at accommodation for IEP and LEP students. Researchers have documented state policies that have encouraged districts to place students in special education or retain them in grade to exempt them from testing, thereby causing average scores to appear higher (a likely problem in other states as well). New policies should ensure this does not continue. Also positive is the state initiating assessments in languages other than English, which should be expanded. Negatively, the graduation exit exam requirement and heavy reliance on multiple-choice can have harmful effects on equity. Data should be reported by demographic categories.

Standard 3: Professional development.

The state has no requirements for teacher competence in classroom assessment. It has not examined classroom, school or district assessment practices or surveyed to determine teacher needs for professional development. Print materials for professional development are available. Professional development opportunities are available through regional centers, and through state trainings, particularly for measurement personnel. Training will be required as part of the phase-in of new assessments. Teachers are involved in writing items for state exams, and in scoring writing samples.

Evaluation. Professional development needs to be expanded and made more systematic for pre- and in-service teachers in conjunction with the new assessments. Teacher involvement in assessment development is positive and should continue; teachers also should be involved in scoring performance tasks on assessments.

Standard 4: Public education, reporting and parents' rights.

This year, the SEA has instituted new school-level report cards which include test data. They are intended to be more accessible than previous reports, partly by the inclusion of graphs and charts. They are available on the Web. The state is translating the report card into Spanish, Chinese and Haitian Creole this year. Russian and perhaps other languages will be included in the future for each school with at least 50 students who speak the language.

Parents have the right to review all tests. Students can take home copies of most state made tests, including the Regents exams, but not tests copyrighted by contractors, such as the Degrees of Reading Power.

Evaluation. Positively, the new report cards should reach more of the public with clearer information, particularly since they are relatively widely translated. The state should survey parents to ensure the report cards are well understood. New York also is very positive in allowing students to take home copies of tests and allowing parents to review tests.

Standard 5: System review and improvement.

FairTest has no data on any planned system review.

Evaluation. Hopefully, the state will carefully and regularly review the new system and its effects, using suggestions from the *Principles*.

New York responded by telephone to the short form of the FairTest survey and forwarded various documents. This report also relied on CCSSO/NCREL, CCSSO and AFT reports. The state responded to a draft descriptive report.

NORTH CAROLINA

Summary evaluation.

North Carolina's assessment program needs a complete overhaul. It relies far too heavily on multiple-choice tests, tests too often, and has a graduation exam. It should reduce the grades tested, drop the graduation requirement, ensure districts do not rely on the tests for grade promotion decisions and implement a performance assessment system based on the state standards. Bias reduction efforts and inclusion are positive. Professional development should be expanded and focused on classroom performance assessment, parental rights should be expanded, and the review process should be revised and strengthened.

Standard 1: Assessment supports important student learning.

North Carolina has content standards and curriculum frameworks for all subject areas, including math, English-language arts, science, social studies, arts, healthful living and foreign languages. The state designs and develops its own tests, which are all aligned with the frameworks. The state reports that its tests use authentic reading material and assess mathematics by focusing on problem solving using real-world information.

The state administers multiple-choice, reading and math end-of-grade tests to all students in grades 3-8. In 1996-97 tests with short and extended-response questions were administered to all students in grades 5 and 8 in reading and math. Writing assessments are administered in grades 4, 7 and 10 (English II essay) and require responses to a single prompt. High school testing in North Carolina encompasses end-of-course multiple-choice tests in algebra I; English I; biology; economic, legal, and political systems; and US history. Local school districts are encouraged to use the results from end-of-course tests (including the English II writing sample) as part of determining students' final course grades.

North Carolina also administers multiple-choice, criterion-referenced competency tests for high school graduation in the areas of reading and mathematics, with a new standard in place for students who are juniors during the 1996-97 school year. Beginning with the class of 2001, students must also pass a multiple-choice and performance test in computer skills. Approximately 40 percent of the students do not meet these requirements by the end of grade 8 and are required to retake the tests in high school until they pass them. In addition, a test in reading and math for all students in grade 10 is scheduled to be implemented during the 1997-98 school year, with school accountability as its main purpose.

North Carolina administers the ITBS, a multiple-choice NRT, to a sample of 3,000 students in grades 5 and 8 in reading, language arts, and math. Local district participation in the NAEP (which also tests a sample of students) is encouraged. Results from these two tests are used to compare the performance of typical students in the state with those in the nation.

A new education reform initiative in North Carolina, the ABC's Plan, started this year. Its major component is a data-driven, school-by-school accountability program. The ABCs Accountability Program holds each school accountable for student progress in basic skills in reading, writing and math. It has two types of performance goals: growth standards are benchmarks set annually to measure a school's progress in improving achievement in reading, math and writing; performance standards annually monitor the percentage of students in the school that perform at or above grade level in reading, writing and math.

In the ABCs program school performance is measured annually against test score expectations. Schools are identified as making exemplary growth if their scores for the year exceed past performance based on a regression formula that takes into account past school scores and the average state growth. Staff in schools making exemplary growth are rewarded with financial bonuses. Schools which fail to meet expected growth and have more than 50 percent of their students achieving below grade level may be identified as low performing. Of those, the 25 lowest performing schools will be assigned technical assistance teams.

Other than the accountability sanctions and high school graduation requirements, there are no other specific consequences attached to the results from the tests. Some LEAs are establishing grade-promotion policies that require students to pass the state tests (rather than just making the tests one part of the grade), in addition to using test scores to inform instruction, and to evaluate local and state-mandated programs.

Evaluation. North Carolina tests far too frequently, relies far too heavily on multiple-choice exams and uses tests for high-stakes purposes. The writing samples and small amounts of constructed-response tasks beginning to be used are positive but still very limited. Using the NRT only on a sampling basis in a few grades is reasonable. The state should reduce the grades tested, shift to predominantly performance assessments that can more fully match high-quality standards and instruction, and drop the high-stakes tests for individuals. For schools, the new accountability program appears also to rely too heavily on testing. Districts must be actively discouraged from using tests as a grade-promotion hurdle.

Standard 2: Assessments are fair.

A bias review committee reflective of the state's demographics reviews test items and other materials. Statistical analyses are also evaluated. Tests use culturally diverse materials in the items and prompts. Reports include subgroup performance by gender and ethnicity.

North Carolina's policy is to include students with special needs in the statewide testing and accountability programs as early as possible. Approximately 15 percent of tested students have IEPs, and approximately 1 percent of tested students are LEP. Modifications and accommodations are available to students with disabilities and to students who are LEP. IEP or LEP students may be exempt from testing as dictated by their IEP or written accommodation plan. LEP students may be exempt for a maximum of two years. Students in both of these categories must pass the competency tests in order to receive a state diploma.

Evaluation. Including demographic data in reports is sound, as is the bias review procedure, though the committee should have authority to delete or modify items. While efforts at inclusion are positive, the over-reliance on multiple-choice testing is a problem for students with IEP and LEP and for equity purposes in general. The high school exit requirement also does not meet this standard.

Standard 3: Professional development.

For licensure, teachers must meet state-established competency standards in assessment. Limited professional development is offered to LEAs on the use of test results. Printed materials about the state testing program; mini-tests in the areas of reading, math, science and social studies (grades 3-8); and an electronic bank with multiple-choice items for reading and math are available for teacher use. Some professional development opportunities in classroom assessment are available upon request. North Carolina no longer surveys educators for their professional development needs because local school districts receive funds and assume responsibility for the professional development of staff.

Evaluation. The over-emphasis on multiple-choice extends to the evaluation of this standard. Too much professional development is focused on this method and on the state tests. Though professional development may be a local responsibility, stronger state guidance and support is warranted, particularly in classroom, performance and portfolio assessment. Teachers are substantially involved in writing items, which is positive, but it would be far better for them to be developing performance tasks.

Standard 4: Public education, reporting and parents' rights.

Test results from all administrations are publicly reported in North Carolina. Parents or schools can challenge test scores, which can lead to rescoring (especially for writing tests and open-ended assessments). Test results are reported in English. All tests are secure.

Evaluation. The state should allow parents to review tests and should survey parents to determine what information they need and whether the reports are understood. Should the state increase the constructed-response portions of its exams, public education will be needed.

Standard 5: System review and improvement.

North Carolina reviews its assessment system annually. An external review team from the University of Alabama recently evaluated the North Carolina Testing Program. The review stated that, "North Carolina's public school testing program is a very ambitious project that was developed and implemented using sound measurement concepts. The testing program meets the current needs as stated in state law and is perceived by both administrators and teachers as a useful tool for instruction and accountability." The state has not systematically evaluated the impact on curriculum and instruction.

Evaluation. Annual review and having outside experts review the system is positive, but evaluation of the impact on curriculum and instruction should be undertaken, as should a study of how well the assessments measure complex and critical thinking in and across the subject areas.

The quotation from the external review clearly reveals a different perspective on what state assessment systems ought to be and do from the perspectives found in the *Principles*. As discussed in the first section of this report, articulating underlying values and goals is fundamental to constructing, as well as evaluating, a state program.

North Carolina responded by telephone to the short version of the FairTest survey and sent various documents. This report also relied on CCSSO/NCREL, CCSSO and AFT reports. The state responded to a draft descriptive report.

NORTH DAKOTA

Summary evaluation.

North Dakota's system needs many major improvements. The state relies on a multiple-choice NRT and administers a "cognitive abilities" norm-referenced test. Both should be dropped. The state tests in four, rather than the recommended three grades. The stakes are relatively low; no required consequences ensue from the test scores. The state is developing new assessments based on its standards which should replace the NRT. Inclusion should be strengthened. Professional development needs substantial improvement. Reporting is currently solid, but public education about the new assessments will be needed. The review process is fairly solid, but can be strengthened in some regards.

Standard 1: Assessment supports important student learning.

North Dakota has voluntary content standards and benchmarks for grades 4, 8 and 12 in English/language arts (ELA), library media, math, science, social studies, arts, business, foreign languages, health and physical education.

The state assessment program consists of testing in grades 3, 6, 8 and 11 in the areas of language arts, math, science, reading and social studies, using both the CTBS/4 NRT and the TCS, a norm-referenced, multiple-choice "cognitive abilities" test. The tests are checked against the curriculum frameworks for alignment, so that what is tested is in the standards. However, the state reports that only about 40 percent of the ELA standards and about 60 per cent of the math standards are measured by the CTBS. The state will be looking for a new NRT that better matches the standards.

An English/language arts project is developing performance assessments for grades 4, 8 and 12. Tasks have been piloted and rubrics developed, and they are being revised. Work has begun on a math performance assessment. For the English/language arts project, alignment to standards is checked through group consensus among educators. Speaking and listening are not included in the English assessment, but will be in the future. Both assessments are expected to be used to meet new federal Title I requirements, but otherwise will be voluntary for districts.

Results of assessments are used for student diagnosis or placement, improvement of curriculum, and program evaluation. They also often become part of the individual student's career portfolio. For schools, assessment results are part of accreditation.

Evaluation. North Dakota relies almost exclusively on a multiple-choice NRT. Rather than adopt a new one, the state should cease to use it. North Dakota appears to be one of the few states still requiring use of a "cognitive abilities" test. This should be dropped for mass use. The development of performance assessments in language and math is a positive step. When developed, they should replace the NRT. Since "cognitive abilities" tests, like IQ tests, have great potential for misuse, the state should be aware of possible misuse in diagnosis and resulting program placement, particularly for minority-group and low-income students.

Standard 2: Assessments are fair.

Bias studies have been conducted for the English language arts assessment. The bias review committee was selected for racial and gender balance. The contractor, CTB/McGraw-Hill, reviews the CTBS and CTS for bias. Disaggregated data are released by race, but not by gender or SES. The contractor provides test practice materials for schools to use in preparing students for the tests.

Nine percent of students tested are classified with an IEP, and 1 percent of students tested are classified with LEP. Students who are either mainstreamed less than 50 percent of the time, or who take tests in a non-standardized manner according to their IEP, can be exempted from the state assessment. Some accommodations are available for IEP, none for LEP, on the NRT. Those who are tested are included in regular reports and separate group reports are released.

Evaluation. We do not know the authority of the bias review committee for ELA, nor how bias review is conducted by the NRT contractor. Inclusion on the NRT appears inadequate. Disaggregated reporting is mostly solid, though data by gender and SES should be released.

Standard 3: Professional development.

North Dakota does not have any pre-service requirements for teachers in assessment, nor does it evaluate teacher competence in assessment. The SEA has surveyed teachers regarding their professional development needs in assessment. The state holds regional test interpretation workshops and uses a workshop evaluation questionnaire for feedback. Voluntary teacher involvement in developing and scoring the new performance assessments in English and math has resulted in professional development. In English, more districts volunteered to participate than had been planned for, creating a shortage of funds.

Evaluation. Professional development needs to be substantially expanded for both pre- and in-service teachers, particularly in classroom and performance assessment. Educator involvement in developing the new assessments is positive, as is conducting the survey.

Standard 4: Public education, reporting and parents' rights.

The SEA has district profiles, including assessment information, and for the last seven years has surveyed educators and some parents regarding several aspects of the state testing program. The feedback has indicated that the report is very understandable for parents and for students above the fifth grade who are given a brief description of the scores.

Parents or students can appeal a score, challenge potentially flawed items, and review the assessment.

Evaluation. The surveys, gathering district information, and the parental rights processes are all very positive. As the new assessments are used, extensive public education about them likely will be necessary.

Standard 5: System review and improvement.

Classroom, school and district assessment practices are studied as part of the North Central Association school improvement programs. A statewide committee, including SEA and LEA staff, conducts ongoing evaluation of the state assessment program. It includes review of the impact of assessment on curriculum and instruction, including the impact of the language arts standards, which are seen has having a positive impact. Information from reviews is used in developing the state assessment program.

Evaluation. The review process seems solid, including the district reviews. As the NRT cannot adequately assess the standards, the effect of the gap between tests and standards needs to be carefully evaluated. The new assessments should be evaluated for validity, for their match to the standards, for their ability to assess critical thinking, and for their impact on curriculum and instruction.

North Dakota responded to the short form of the FairTest survey. This report also used CCSSO/NCREL, CCSSO and AFT reports. The state responded to a draft descriptive report.

OHIO

Summary evaluation.

Though Ohio did not respond to the FairTest survey, it appears from other data sources that the state assessment program needs many major improvements. The tests are mostly multiple-choice, aside from the writing sample, and the state has a mandatory high school exit exam. The burden is moderately heavy, with five subjects tested in each of four grades. The addition of constructed-response items is positive, but they are not a major part of the exams. Bias review information was not available. Inclusion is substantial, but for LEP it is not clear that the assessments are always appropriate. Professional development does not appear to be adequate, though little information was available. We received little information on public reporting, and none on parent rights or on review of the system.

Standard 1: Assessment supports important student learning.

Ohio has standards in English language arts, math, social studies and science. They are under development for foreign languages, the arts, and health and physical science. The state has content standards. Performance standards exist for designated grades in reading, writing, math, citizenship and science. The state claims its assessments are aligned with the standards.

The state assessment program consists of proficiency testing at grades 4, 6, 9 and 12 in the areas of reading, writing, mathematics, science, and citizenship. All tests are criterion-referenced and primarily multiple-choice, with some gridded-in and constructed-response items on the grade 12 math test and some short- and extended-response items on the various grade 4 and 6 tests. Other alternative assessments are in development, but their use has not been set.

The writing assessment is based on responses to SEA provided prompts twice a year. Each time, students are given 150 minutes to produce a writing sample. Schools are required to give students more time (up to 2.5 hours) as needed.

The grade 9 results are used as a high school exit requirement for a regular diploma. All students must pass it, though students who do not can submit sufficiently high scores on the ACT or SAT college admissions tests as an alternative.

Results of the assessments are used for improvement of curriculum and instruction, accountability and school performance reporting. The grade 4 results are used for remedial intervention in grade 5 for students who do poorly, the grade 9 results for individual improvement instructional plans as needed, and the grade 12 results for honors high school diploma. Results of the grades 4 and 9 exams make schools eligible for competitive grant funds targeted for school improvement.

Evaluation. With high stakes and mostly multiple-choice, Ohio needs major improvement. While the state reports its assessments match its standards, with mostly multiple-choice it is likely that significant areas of the standards are not assessed. Extended time for response on the writing sample is positive, as is the inclusion of constructed-response items. However, writing to a prompt is a limited measure of writing. While use of SAT or ACT test scores are an option for high school graduation, this is an inappropriate use of those exams.

Standard 2: Assessments are fair.

Print materials are available to prepare students for the exams. Data disaggregated by race and gender are released at the school, district and state levels.

All students must be tested unless specifically exempted by their IEP. For IEP, some accommodations in test format and procedure, but not content, are available. All LEP students are required to be tested, but may have a "translation dictionary," extra time and other limited accommodations on some of the tests. LEP seniors who have not passed all parts of the grade 9 exam may have an oral administration, except for writing, and may have the assistance of an interpreter for math and citizenship. For IEP students, inclusion or disaggregation in reports depends on accommodations they have had. LEP students' results are included and not disaggregated.

Evaluation. Disaggregated reporting also should be done by LEP and SES students. Inclusion will need improvement for IEP students. Requiring all LEP students to take the regular exams, even with some accommodations, is probably inappropriate at times; alternate exams are probably needed.

Standard 3: Professional development.

Print and video materials are available to educators for professional development.

Evaluation. Ohio reported no additional professional development for the state assessments, and we have no data on any other professional development in assessment.

Standard 4: Public education, reporting and parents' rights.

Print materials are available to parents, policymakers and the public for educational and information purposes.

Evaluation. No information was available about parent rights or public education.

Standard 5: System review and improvement.

No information was available.

Ohio declined to participate in the survey. This report used two years of CCSSO/NCREL reports and the AFT report.

OKLAHOMA

Summary evaluation.

Oklahoma's assessment program needs many major improvements and perhaps a complete overhaul. It relies far too heavily on multiple-choice, uses an NRT on all students in two grades, and shows weak performance on most of the other standards. While bias review is solid, accommodations for IEP and LEP students are limited. Professional development is inadequate. Parental rights and public education are acceptable given the narrowness of the testing program. Reviews have not been adequate, but a stronger review is planned.

Standard 1: Assessment supports important student learning.

Oklahoma has content standards, the Priority Academic Student Skills (PASS), to which state assessments are aligned. PASS standards exist in language arts, math, science, social studies, the arts, languages (including Native American, foreign and American Sign), health/safety and physical education, instructional technology, information skills and technology education for grades 6-10.

Oklahoma's assessment program consists of two tests. One is a commercial NRT (the ITBS) for grades 3 and 7 in the areas of math, language arts, science, social studies, and reading.

The second is the Oklahoma Criterion-Referenced Tests (CRT). It contains multiple-choice items in science, math, reading, history and government, plus a writing sample, for grades 5, 8 and 11. Performance standards based on the PASS standards are set for each content area of the CRT. The SEA is developing tests in geography and culture and arts for the same grades, plus an 11th grade Oklahoma history exam. All students are tested and all students see the same items. Tests are developed by a contractor. In writing, students are given 50 minutes to write to a prompt; scoring is by a commercial company.

The tests are intended to guide curriculum and instruction. The SEA recognizes that not all aspects of the standards can be measured using multiple-choice items. The SEA explained, "Hopefully, classroom teachers address many of these skills using non-traditional assessment approaches." Also, the effort has been made to include "higher order cognitive demands" in the items. The NRT "does not fully overlay the state's core curriculum," but was selected as the NRT most aligned to the state's curriculum.

Input from teachers and administrators is used to select which NRT best aligns to state curriculum and possesses other desired qualities. Educators have participated in the development of the CRT.

Assessment results are used for student diagnosis or placement, curriculum improvement and program evaluation. The CRT "results indicate whether student has met the Satisfactory Performance Standard Set for the specific grade and content area." If a student does not attain a satisfactory level, remediation must be offered in the subject and the student retakes the test the following year. Accountability for schools includes school performance reporting with possible consequences (for the NRT component only) such as warnings and probation. Accreditation loss, takeover and dissolution can occur, with NRT results as one element in such decisions.

Evaluation. The state needs to rethink its assessment program, moving away from reliance on multiple-choice CRTs and NRTs and toward performance assessments that can more fully assess to high-quality standards. Stakes are not too high, as test scores are only one factor in program evaluation, though the NRT used for that is not based on the state's standards. So long as the CRT and NRT are intended to guide curriculum and instruction, assessment will not adequately support important learning.

Standard 2: Assessments are fair.

SEA has content and statistical review of the CRT for bias, including pretesting items. A bias review committee, with "cross-sectional representation," can "vote out" items. NRT items are pretested for bias by the contractor. The state provides a practice test for the grade 3 ITBS, and a booklet is given to teachers in grade 3 and 7 to help them align instruction to the test. Student study guides for the CRT include the standards and practice tests. Disaggregated data are not reported for sub-populations.

13 percent of students tested are classified as having IEPs and 3 percent are LEP. Exemptions from testing are allowed, and limited accommodations are available for students with IEPs. Both are included in regular reports.

Evaluation. Bias review seems solid for the CRT, but reporting does not include data for sub-populations. Informing students about the tests seems sufficient. Accommodations are limited and alternatives are not available for students for whom accommodations are not sufficient. As fairness involves use of multiple methods of assessment, Oklahoma does not meet that aspect of the fairness standard.

Standard 3: Professional development.

The state has required in-service training for building and district test coordinators (voluntary for administrators) on the administration of standardized tests. Workshops on classroom assessments and traditional tests are offered. The state does not examine teacher competence or needs for training in assessment, or survey district, school or classroom assessment practices. Post-test in-service workshops on the NRT and the CRT are conducted to help teachers ferret out information from the score reports, to use in curriculum planning.

Evaluation. Professional development is not adequate, being neither comprehensive nor systematic. Most professional development is geared toward using the multiple-choice tests and not helping develop and use performance and classroom assessments.

Standard 4: Community participation and reporting to the public.

Parent study guides for the CRT include the standards and practice tests. Parents can preview or review the CRT and the NRT under controlled conditions. Parents can request an exemption from the NRT. The state has not surveyed the public for information it wants, but says it works to make reports user-friendly. CRT results are reported in August, in English only. NRT results are returned in 6 weeks and publicly released in 3 months.

Evaluation. Given the limitations of the testing program, parent education seems acceptable and parents rights are reasonable. Should the system make recommended changes, it will need to do more extensive public education.

Standard 5: System review and improvement.

The state system is not reviewed, and technical studies have not been conducted on the CRT, but the state views the participation of educators in developing the CRT as a form of review. The tests are intended to guide curriculum and instruction. The state plans to study the consequences of the CRT on curriculum and instruction, tracking, etc. The state reports that rising scores on both assessments may be evidence of a positive curriculum impact.

Evaluation. Reviews have been inadequate, lacking technical and consequential studies. While the state now plans to review the impact of the CRT, it is not reviewing the impact of the NRT. Rising scores may be a sign of real learning or only inflated test scores. It is not clear the extent to which a review will analyze the ability of the tests to measure cognitively complex work or critical thinking in any tested subject. Thus, while the planned review is welcome, it is not yet sufficient.

Oklahoma responded to the full FairTest survey. This report also used CCSSO/NCREL, CCSSO and AFT reports. The state responded to a draft descriptive review.

OREGON

Summary evaluation.

Oregon's assessment system needs some significant improvements. It is not clear whether pending changes will make the system better or worse. Currently, the state relies too heavily on multiple-choice testing and tests some subjects in four, not the recommended three grades. The exams and the proposed uses of them for certifications and college admissions could produce a program weighted far too heavily toward standardized tests. The state's bias reduction and efforts to include IEP and LEP students are solid, but professional development needs major strengthening. Public education is currently adequate, but will need strengthening, as will parental rights. The evaluation process may be sufficient, but critical areas need attention to ensure this is the case. The state should plan to review local assessment practices.

Standard 1: Assessment supports important student learning.

Oregon has content standards for K-12 in English language arts, math, science, social studies, arts and second languages. Performance standards in English and math also have been adopted, for grades K-10. The state claims that assessments are aligned with the previous common curriculum and will be aligned with new standards for Certificates of Mastery.

The Oregon state assessment program includes criterion-referenced Reading/Literature and Mathematics Assessments in grades 3, 5, 8 and 10. These are primarily multiple-choice tests with constructed-response items in math in grades 5, 8 and 10. All students are tested, and multiple forms are used. Students are also assessed in writing in grades 3 or 5 and 8 or 10. The writing assessment is based on samples produced in response to SEA provided prompts. All students are tested, but different students see different prompts. Students are given approximately 45 minutes per day over three days to produce a sample. Scoring is done by teachers and others with a BA in English under the SEA's direction.

A multiple-choice science test is planned for 1997-98 for grades 5, 8 and 10, with a state exam in social studies to follow. The state is developing Certificates of Initial Mastery and Advanced Mastery, to be earned based on passing state assessments and classroom assignments at about grades 10 and 12, respectively. These will be separate from the diploma. It also is developing standards and assessments, based on the K-12 standards, for use in admissions to state colleges. The real consequences attached to the Certificates and the relationship among these various assessments is not yet defined.

Alignment between the standards and the assessments is determined by grade-level expert teachers who review test items against state curriculum benchmarks. A team of national experts also evaluated the alignment. It is recognized, however, that not all aspects of the standards are assessed.

Results of assessments are used for curriculum improvement and program evaluation and for reporting on schools and individual students. No high-stakes consequences are attached to assessment results for schools.

Evaluation. The state tests rely too heavily on multiple-choice. New exams will perpetuate this problem. Currently, the stakes are not high. Of potentially major concern will be the assessments used for the Certificates (if they come to have major consequences) and entry to college. They should not be single assessments, should be classroom-based, and should be primarily constructed-response and performance assessments.

Standard 2: Assessments are fair.

Sample tests are provided to schools. Spanish-language versions of tests have been developed. Reading selections are multicultural. The use of multiple measures is intended in part to address different learning styles. The grade 3 tests have been reviewed for developmental appropriateness. Public reports provide data by gender, race and SES.

A bias review committee can recommend changes; all major ethnic/racial groups in Oregon are represented. Items are also statistically analyzed for bias.

Nearly 11 percent of students tested have an IEP and 1 percent of students tested are LEP. Oregon's policy on LEP and IEP students is to be as inclusive as possible; however, decisions are made locally as to whether a student takes the test or an adapted version, or is provided one of the available accommodations, or is not tested.

Evaluation. Bias review and disaggregated reporting appear solid. While the state has reviewed for developmental appropriateness and uses multiple methods of measurement to address different learning styles, the over-reliance on multiple-choice does not meet this standard. Policies for inclusion of IEP and LEP students appear mostly acceptable, but the state needs to monitor actual district practice to ensure full inclusion.

Standard 3: Professional development.

The state has no pre-service requirements in assessments, and it does not evaluate teacher competence in assessment or survey educators about their professional development needs. Printed and TV materials are available to educators for professional development purposes, along with workshops and telecourses. The state has regional scoring centers for writing and math problem-solving where 1,000 teachers annually score assessments.

Evaluation. Professional development for pre-and in-service teachers must be strengthened and expanded, particularly in relation to classroom assessments and to the new Certificates and college admissions process. The scoring centers are a positive development.

Standard 4: Public education, reporting and parents' rights.

Printed materials are available to parents, policymakers and the public for information purposes. Public reports are produced only in English. Students who took the test in Spanish and their parents receive individual reports in Spanish. Parents have not been surveyed as to what information they want or whether the reports are understandable. Parents can review tests after administration, but no procedures exist for challenging scores or possibly flawed items.

Evaluation. Public education will need to be strengthened as new assessments and new uses of assessments are implemented. Reporting to Spanish-speaking students and parents in Spanish is positive, but the public reports also should be translated. Surveys should be used and parent and student rights strengthened, particularly as assessments are used for more purposes.

Standard 5: System review and improvement.

The state has not evaluated assessment practices at the classroom, school or district levels. Since changes are being made, the state system is now reviewed annually by the SEA and independent reviewers. The assessment is supposed to guide curriculum and instruction, and its impact is part of the program evaluation. The evaluation will be used to guide revisions in assessment. Technical studies are being done. To date, whether or how well the assessments elicit critical thinking has not been evaluated.

Evaluation. The plans for review and evaluation mostly are solid. How well the tests actually assess to the standards, particularly where critical thinking is part of the standards, must be studied carefully. Review of district practices should be done, particularly since local assessments are planned to be part of the new Certificates and college admissions processes. The impact of the Certificates on school completion also needs to be studied, as does their impact on tracking and placement issues.

Oregon responded to the short form of the FairTest survey. This report also relied on CCSSO/NCREL, CCSSO and AFT reports. The state responded to a draft description.

PENNSYLVANIA

Summary evaluation.

Pennsylvania's assessment system needs some significant improvements, primarily a shift away from mostly multiple-choice testing toward performance assessments. Positively, the state does not have higher stakes than public reporting and does not rely on NRTs. The state is emphasizing district standards and assessments, but the SEA should support districts in implementing performance assessments. Bias review and inclusion of students with IEP or LEP seem mostly solid, though reporting by sub-populations should be done. Professional development should be expanded and made more systematic, either directly by the state or with the state supporting the districts. Reporting and public education are currently adequate. The state's review system is not in place as new assessments are being developed, but such a system is needed and should be comprehensive.

Standard 1: Assessment supports important student learning.

Pennsylvania has standards in the form of *Learning Outcomes,* which have been subject to substantial controversy. An advisory committee of educators (with SEA guidance) is developing new content/performance standards and assessment frameworks in reading, writing and math (science is on hold, and others may follow) that are intended to replace the outcomes. The state currently has non-mandatory curriculum frameworks in math, based on the *Outcomes*, and English, not based on the *Outcomes*. The state claims its assessments are aligned to the *Outcomes* but that "alignment has not been checked in a formal study involving content committees independent of the advisory committees that initially evaluated alignment." The assessments are being revised to match the new Pennsylvania standards.

The state assessment program consists of the criterion-referenced Reading and Mathematics Assessment for grades 5, 8 and 11. The test is mostly multiple-choice but includes some constructed-response items. A writing assessment for grades 6 and 9 uses writing samples with SEA and teacher-provided prompts. It is mandated to be given to one-third of the schools each year and is voluntary for the remaining districts or schools; about 90 percent of the districts actually participate annually. All students in the included schools are tested. Students are given 80 minutes over two days to produce a sample. Scoring uses a rubric developed by teachers with SEA guidance and scored by state teachers in conjunction with a commercial company.

Districts are asked to describe their plans to assess in nine academic areas. District-developed standards will be used in writing the new state standards.

Results of assessments are used for curriculum improvement, program evaluation and school performance reporting. No high-stakes consequences result from the assessment.

Evaluation. Pennsylvania should shift from mostly-multiple-choice to mostly-performance or constructed-response assessment. The districts should also be encouraged and supported in developing or adopting performance assessments and portfolios. The state seems to be regularly changing its outcomes or standards, making alignment more difficult and creating periods in which the tests will not match the new standards. The absence of high stakes and the light testing load are positive.

Standard 2: Assessments are fair.

The state has had ongoing bias review, mostly statistical analysis. It now has a "Fairness Review Task Force" which plans to look at all aspects and forms of assessment. It will be able to remove or revise items and will attempt to improve tests. The state does not now report data by sub-populations.

Eight percent of students tested have an IEP and one percent are LEP. Most students with IEP or LEP are included in the assessment. A range of accommodations are available. Some are excused if they cannot perform adequately on the assessment. The results of all tested students are included in regular reports.

Evaluation. With the addition of the fairness task force, bias review seems adequate. Inclusion seems most positive, but development of alternative assessment for those for whom accommodations cannot be used is recommended.

Standard 3: Professional development.

Assessment Handbooks are provided to teachers and administrators for professional development. The state has portfolio projects and other assessment activities which districts can participate in and which provide professional development. Some information about professional development needs comes from local planning reports, which often include assessment training needs. Requirements for pre-service teachers were not reported.

Evaluation. Professional development seems on the right track, but likely needs to be expanded and made more systematic, preferably together with the greater use of performance assessments by the state and districts. More systematic gathering of information regarding training needs, and evaluation of teacher competence in assessment, are recommended.

Standard 4: Public education, reporting and parent rights.

State *Assessment Reports* are provided to educators, parents and policymakers for informational purposes, in English only. One *Report* is an in-depth study of the data which is provided to districts for use in their reports. A revised *School Profile*, intended to be made public via CD ROM this year for the first time, will include demographic data and some score data at school, district and state levels. Parents and local educators provided input as to what the profiles should contain. A booklet about the tests is available to parents before the exam. Parents can opt their children out of state assessments; the state reports that opting out is probably increasing slightly.

Evaluation. As new assessments are implemented, public education should be enhanced. Meanwhile, public education and reporting seem fairly solid.

Standard 5: System review and improvement.

Assessments are being reviewed in order to revise them to meet new standards. Issues such as whether the assessment adequately matches the standards or assesses higher order thinking in domains will probably be included in future reviews of the state's assessments. The state does not directly study the impact of assessment on curriculum and instruction, but some indirect information is available from local reports. Local reports also provide information on what tests are used.

Evaluation. Reviews need to be expanded to cover technical and consequential issues, including alignment, impact on curriculum and instruction, and assessment of cognitively complex subject matter.

Pennsylvania responded by telephone to the short form of the FairTest survey. This report also used CCSSO/NCREL, CCSSO and AFT reports. The state responded to a draft description.

RHODE ISLAND

Summary evaluation.

The Rhode Island assessment program needs some significant improvements, notably reducing or eliminating the NRT, but it is developing in a positive direction. The use of performance assessments in its new exams is positive. Fairness efforts are clearly positive, professional development and community education have positive features but need expansion, while evaluation needs much improvement.

Standard 1: Assessment supports important student learning.

Rhode Island has developed curriculum frameworks in language arts, math, science and health. Public involvement in their design included a development committee and focus groups; information sessions were held after the frameworks were developed.

The Rhode Island state assessment program includes an NRT, the MAT/7, for grades 4, 8 and 10, using only the subtests in reading comprehension and mathematics concepts and problem-solving. Criterion-referenced exams are aligned with the curriculum frameworks. They include: a writing assessment for grades 4, 8 and 10; a math performance assessment in grades 4 (SEA developed), 8 and 10 (New Standards Reference Exam); and a health performance assessment in grades 4 and 8. All students are assessed and all see the same items. For writing, students have 45 minutes over two days to respond to an SEA prompt.

Results of assessments are used for general accountability, student diagnosis or placement, curriculum improvement, program evaluation, and school performance reporting. There are no high-stakes consequences from the results of assessments for either schools or students, though school accountability procedures may expand. The state plans to add performance assessments in reading and science. Work is underway on a "Certificate of Initial Mastery" which students would earn through a variety of assessments.

Evaluation. By using only a portion of the NRT, Rhode Island helps minimize its impact, but shifting to sampling would be preferable. Positively, the math assessments are constructed-response and criterion-referenced. The testing weight does fall heavily on a few grades and could perhaps be spread out. The stakes are relatively low. It will be essential to ensure that, in developing a Certificate of Initial Mastery program, the state does not make any one assessment a gatekeeper, and that it relies on performance assessments and portfolios.

Standard 2: Assessments are fair.

A bias review committee with a wide variety of racial/ethnic participation, including minority over-representation, reviews state-made assessments and has authority to remove or alter items. Items are analyzed before administration and sometimes after, if problems have been identified during the scoring process. Efforts are made during item construction to respond to different learning styles.

Eight percent of students tested have an IEP, and 5 percent of students tested are LEP. On the NRT, IEP students who spend more than 50 percent of their time in a self-contained classroom and LEP students with less than two years in the US may be exempted. On the CRTs, the intent is to test all students (most are), and extensive accommodations, including translation to Spanish (except writing, which requires responses in English), are available. (In 1995-96, tests were also translated to Portuguese, Khmer and Lao; budget restrictions prevented this in 1996-97.)

Evaluation. Bias reduction efforts and inclusion of IEP/LEP students via accommodations on the CRT are quite positive. It is unfortunate that funding constraints have reduced translations.

Standard 3: Professional development.

The state has no professional development requirements in assessment. The SEA provides "extensive" training on performance and classroom assessments, while recognizing it has limited capacity to train all teachers. Teacher competence in assessment is not evaluated, nor are educators routinely surveyed about their professional development needs.

Evaluation. Strengthening the program by building on the current training component is important, particularly as the state implements more performance assessments.

Standard 4: Public education, reporting and parents' rights.

Written information is sent to teachers to prepare for tests. Reports are available only in English, but some information about standards and assessments is printed in other languages. "Interpretation guides" that explain the assessments are publicly available and also are distributed to teachers in grades 4, 8 and 10. The SEA has not surveyed parents or the public about what information they want or whether state reports are understood.

Parents can exempt their children from testing. Students cannot appeal scores as the stakes are low. They can challenge items, but this has not happened. Parents cannot generally review assessments since they are secure and reused.

Evaluation. Reporting efforts appear adequate and parental rights acceptable, except that secure review should be allowed. Further public education as more performance assessments are implemented will be necessary.

Standard 5: System review and improvement.

The SEA plans to survey district assessment practices in 1997-98. The state does not have a formal review process for its assessments, but the SEA informally reviews the program to determine what should be offered. The state cites anecdotal evidence that the performance assessments are changing curriculum for the better, but studies of the impact of assessment on curriculum and instruction have not been done. Alignment with standards has been evaluated by the SEA with staff from New Standards, a state contractor. Not all aspects of the standards are assessed, and the state has not evaluated the exams to determine the extent to which they assess critical thinking. Technical studies have not been conducted.

Evaluation. A substantially stronger evaluation system is warranted.

Rhode Island sent information and responded to the short form of the FairTest survey and a draft description. This report also relied on CCSSO/NCREL, NCREL and AFT reports.

SOUTH CAROLINA

Summary evaluation.

South Carolina's assessment program needs a complete overhaul. The new program now being implemented, which will be based on new state standards, is not a sufficient change. In its old and new systems, the state relies too heavily on multiple-choice testing. Currently, it relies too heavily on an NRT, and it may continue to do so. The state should build some initial work in performance assessments as a starting point for redesigning the system to a largely performance-assessment system. It also should either eliminate or reduce to sampling the NRT and drop the graduation requirement. Bias review and policies for inclusion of IEP and LEP student appear adequate, but as much of this is under LEA control, it is hard to be sure of the extent of inclusion. Current and planned professional development is not adequate for high quality assessment practice by educators. Public education efforts appear solid. Information on review processes was not available.

Standard 1: Assessment supports important student learning.

South Carolina has completed curriculum frameworks along with content and achievement standards in math, English language arts and science. Frameworks with content standards have been developed in foreign languages, and visual and performing arts. They are in development for social studies, health and safety, and physical education. These documents are developed statewide by committees of educators and business people, with broad public review and input.

Current tests are not aligned to the standards. The Basic Skills Assessment Program (BSAP) consists mostly of criterion-referenced, basic skills, multiple-choice tests for students in grades 3 and 8 in the areas of science, reading and math; science at grade 6; and math and reading at grade 10. An off-the-shelf NRT (MAT-7) assesses all students in grades 4, 5, 7, 9 and 11 in the areas of language, reading, and math. The state requires a readiness test in grade 1, a modified version of the individualized, teacher-administered Cognitive Skills Assessment Battery, intended to help teachers identify student capability in reading and math and thus guide instruction.

Grades 6, 8 and 10 are also tested in writing through SEA-provided prompts. Students are given up to the length of school day to produce a writing sample with revisions. Scoring is done by a commercial company using rubrics developed by a committee of teachers.

Results of these assessments are used for student diagnosis or placement, curriculum improvement and program evaluation. The tests comprise 25 percent of the student promotion criteria in the grades tested. The Basic Skills component in grade 10 is used as a graduation requirement. Accountability for schools includes awards or recognition and performance reporting. High-stakes consequences include funding gains, exemption from regulations (Basic Skills only), warnings and probation/watch lists.

Assessments are being developed to meet the new standards in at least grades 3, 6, 8 and 10 (a revised high school exit exam). The legislature is considering whether to require testing in all grades. The exams will be mandatory in language arts, math and science, but may not be in other subject areas. The state intends to have a criterion-referenced system based on the standards, and not a norm-referenced system. The state is therefore considering how to obtain national comparative data without using full NRTs, beginning in 1998 (it may embed items from an NRT or use a short-form NRT). The methodology will be primarily multiple-choice, with some short or longer constructed-response items. Current consequences will continue with the new assessments. Subscores reporting student performance in relation to the standards will be provided to schools to help guide improvement.

Over the past five years, teachers in selected schools have been developing performance-based assessments as a pilot program to improve instruction and learning. These are disseminated statewide. Next year, the state also will make training available to K-3 teachers in the use of the Work Sampling System (WSS) or a South Carolina version of the Primary Learning Assessment System (PLAS). This will be a voluntary program.

Evaluation. The state relies far too heavily on multiple-choice exams and on the NRT. The state should either drop the NRT or use it only on a sampling basis in a few grades. The new program will test in at least four grades, one more than the recommended three. The new tests will be disproportionately multiple-choice and should be revised. The graduation requirement should be dropped. The combination of high stakes for schools and individuals with mostly multiple-choice tests is likely to heavily influence curriculum and instruction in a narrow direction. Making performance assessments, the WSS and the PLAS available are positive steps. They should be made more important parts of the state assessment system while the traditional tests are deemphasized.

Standard 2: Assessments are fair.

Currently, each year's new writing prompts are reviewed for bias. In the new assessments, all items will be reviewed, and items determined by the committee to be flawed will be discarded. Items will be previewed by teachers under secure conditions, including pre-testing on students, as writing prompts now are. Print and video materials are available to students for explanatory information. The state does include reporting by SES and race, but not gender.

LEAs decide whether IEP or LEP students will be included in the assessment and which, if any, of the extensive available accommodations will be allowed, including on the high school exit exam. Alternate scoring scales are available for the writing test. For IEP students, the reading and math tests can be administered orally. Results for all students tested are included in regular reports. IEP and LEP students must pass the grade 10 test to earn a diploma.

Evaluation. Bias review procedures planned for the new assessments appear to be adequate. As LEA's determine inclusion, it is difficult to know the extent of participation by IEP or LEP students. Accommodations appear adequate. The state should also report data by gender. The graduation requirement and the over-reliance on multiple-choice do not meet this standard.

Standard 3: Professional development.

Print and video materials for professional development are available to educators. All the state's professional development efforts focus on the standards, and many will involve using assessments based on the standards. Schools are surveyed about their use of the standards.

Evaluation. Professional development is focused on state standards and the inadequate new state assessments. The support in performance assessment and the planned trainings in using the WSS and PLAS are positive. They should be built on to provide systematic professional development to pre- and in-service teachers in classroom assessment. Educators should be surveyed about their professional development needs.

Standard 4: Public education, reporting and parents' rights.

Print and video materials are available to parents and policymakers. A statewide newspaper supplement was recently disseminated to parents explaining the new system.

Evaluation. Public education appears adequate. We did not receive information about reporting.

Standard 5: System review and improvement.

We did not receive information about the review processes in South Carolina, and a new system is being introduced. A strong review system should be part of the new assessment system. The impact of the readiness and writing tests, as well as the new exams, on curriculum, instruction and placement should be reviewed.

South Carolina responded by telephone to parts of the short FairTest survey. This report also used CCSSO/NCREL, CCSSO and AFT reports. The state replied to a draft description.

SOUTH DAKOTA

Summary evaluation.

South Dakota's assessment system needs many major improvements. It relies entirely on multiple-choice NRTs. Unfortunately, legislative action in 1997 has made the situation worse by increasing the amount of testing without improving the instruments used. The legislature should appropriate funding for the state to develop or adopt performance assessments that can assess to the state's standards. It also should not assess in grade 2, as it now plans to do. Issues of equity and adequate system review must also be considered, and professional development expanded. Positively, the stakes remain relatively low.

Standard 1: Assessment supports important student learning.

South Dakota has adopted content standards in math, science, reading, writing, social studies and the arts. Districts now must either adopt the state standards or develop their own equally challenging standards.

The state's current assessment program includes off-the-shelf, multiple-choice NRTs for grades 4, 8 and 11 in the areas of English, math, science and social studies; and an off-the-shelf, multiple choice NRT for students in grade 9 on aptitudes and career interest (the Career Planning Program from ACT). Private and home-schooled children are required to take the NRT. Pretesting in grade 4 is done for practice on the grade 4 NRT.

The new legislation calls for testing in grades 2, 4, 8 and 11 in reading, math, science and social studies. There will be a writing exam in grades 5 and 9. As no funding was appropriated for new assessments, the state will use an NRT that is supposed to be aligned with the standards. The SEA recognizes, however, that "you can't test standards with an NRT." The SEA hopes to obtain money to develop its own assessments.

Results of assessments are used for curriculum improvement, career planning, and as a "barometer of local scores to state average." There are no specified consequences for students, schools or staff from assessment results, except that failing to administer the test could jeopardize school accreditation. The program is deliberately low stakes.

Evaluation. Positively, the stakes are low, and introducing a writing assessment will help balance the multiple-choice approach of the NRTs. However, the state is also regressing by initiating large-scale testing in grade 2, which is not developmentally appropriate. As recognized, it will not be assessing its own standards.

Standard 2: Assessments are fair.

Items are pre- and posttested and analyzed by the publisher for bias. Eight percent of students tested have an IEP, and 1 percent of students tested are classified as LEP. Both may be excluded from testing. A few accommodations (Braille, small group administration) exist for IEP on the NRT. Gender and race data are reported at the school level on the current NRTs.

Evaluation. Accommodations for LEP and IEP are not sufficient, and multiple methods of assessment are not used.

Standard 3: Professional development.

The SEA offers professional development in classroom assessment approaches, state assessment programs, and the use of test results, to teachers, school administrators and other school personnel. Professional development in performance assessments is available throughout the state for teachers in a specific program or project. Workshops on pretests and posttests are available each year. The SEA does not survey teachers for needs or evaluate their assessment competence.

Evaluation. The approach and range of content for professional development is positive, though expansion seems warranted to ensure competence in classroom and large-scale assessment by all teachers. Teachers should be involved in both designing and scoring the writing assessments.

Standard 4: Public education, reporting and parents' rights.

NRT results are reported in the spring of the test year to students, parents, and schools and in the fall after testing to the public, in English only. The state provides guidance on the use of results to teachers, school administrators, psychologists/counselors, district administrators, and the general public. The SEA has not surveyed parents as to what assessment information they desire or whether the reports are understandable.

Evaluation. Given the nature of the testing, reporting is probably adequate.

Standard 5: System review and improvement.

A team of SEA staff, consultants, teachers and administrators reviews the state assessment yearly. No studies have been done regarding the consequences of the test. The SEA has not evaluated district, school or classroom assessment practices.

Evaluation. As the state is mandating standards to districts, the impact of NRTs that fail to assess to the standards should be carefully evaluated. Particular emphasis should be paid to the impact of the new grade 2 exam.

South Dakota responded to the full FairTest survey. This report also used CCSSO/NCREL, CCSSO and AFT reports. The state replied to a draft description.

TENNESSEE

Summary evaluation.

Tennessee's assessment system needs a complete overhaul. It relies almost entirely on multiple-choice items, uses norm-referenced testing, and has a high school exit test. The state tests young children with a multiple-choice NRT, and it tests in too many grades. Bias review may not be satisfactory, and many students with IEPs are not assessed. Some aspects of professional development are progressing, but others are not addressed. Parental rights need expansion in some areas. Public education may be adequate, but the state reporting system is very complex. Reviews have not been adequate and need major strengthening. A few changes made in the 1997 legislative session have not improved the situation.

Standard 1: Assessment supports important student learning.

Tennessee has developed frameworks in all areas if the curriculum. These are periodically revised in conjunction with textbook adoption and now also to align them with national standards.

The Tennessee Comprehensive Assessment Program (TCAP) has three components. The Achievement Test is a customized edition of the CTBS/4 multiple-choice test producing norm- and criterion-referenced results in math, reading, language arts, science and social studies. It has been administered to public school students in grades 2 - 8 and to home-schooled students in grades 2, 5 and 7. Items are written by teachers, content specialists and outside experts to a design by the test publisher and SEA assessment staff. The legislature recently dropped using TCAP in grade two, but ordered implementation of basic skills tests in grades 1 and 2.

The criterion-referenced, multiple-choice Competency Test in language arts and math is given to public school students beginning in grade 9. It is offered four times annually, and students must pass it to receive a regular diploma. Teachers, administrators, SEA assessment staff and content specialists are involved in all areas of design and implementation.

Writing is assessed in grades 4, 8 and 11 using responses to SEA provided prompts. Teachers, administrators, content specialists and outside experts are involved in all phases, from design to scoring. Students in grades 4 and 8 are given 35 minutes and students in grade 11 are given 25 minutes to produce a writing sample on demand.

Tennessee has begun to implement high school subject matter tests in pre-algebra, algebra I and II, geometry and math for technology. These are customized, multiple-choice tests developed by teachers and the contractor.

The tests are intended to be aligned to state standards (the Competency Test to the grade 8 standards). Alignment is determined by having educators in content areas write items.

How results of assessments are used vary with each component. All results may be used for instruction improvement or curriculum development. The achievement and competency tests are not intended to guide curriculum or instruction, except to the extent necessary to cover competency test objectives adequately.

Achievement Test results are also used in program evaluation, warnings, probation and takeover, and staff accountability. Competency Test results are used for student diagnosis, awarding a regular high school diploma and for school performance reporting. The high school subject matter tests will provide diagnostic data for districts, schools and teachers. Tennessee has developed a complex "value added" approach to try to measure the score gains attributable to specific schools and to hold schools accountable for adequate score gains.

Evaluation. Tennessee has developed an assessment system organized around high-stakes multiple-choice tests. While the tests are intended to be aligned to the standards, multiple-choice is not adequate for measuring to high quality standards and alignment cannot be determined simply by having teachers write items in given content areas. The time allotted for writing is too short for an adequate assessment of writing. While the state may say that the tests are not intended to guide instruction, it is very likely that the high-stakes tests do have a substantial classroom impact. Controlling schooling through narrow tests runs directly counter to the *Principles* and these Standards. Thus, Tennessee should restructure its assessment system.

Standard 2: Assessments are fair.

Items on all tests are reviewed pre- and post-administration for bias. The contractor for the Achievement tests has a review committee. Bias review for the Competency Test is conducted by the state Testing and Evaluation Center at the University of Tennessee, which can delete items deemed biased. Item selection attempts to take account of the cultural variety in the state. For the writing test, the state's cultural diversity is considered in prompt selection, but no bias review committee exists.

The SEA provides material to familiarize students with test format and typical content. Explanatory information is provided to teachers and administrators, and to students on the competency test. Test reports are only in English and do not report data disaggregated by sub-populations.

About 19 percent of the state's students have an IEP, while 8 percent of students tested have an IEP. Less than 1 percent of the state's students are LEP. A somewhat limited range of accommodations is available. Both IEP and LEP students may be exempted from any of the tests, but must pass the tests to obtain a regular diploma.

Evaluation. Bias review may not be sufficient in that it is not done for the writing prompts and it is contracted out for the competency and achievement tests. Reporting should include data by demographic groups. It appears that many students with IEPs are not tested, though the state's exams may not be appropriate for many students. Heavy reliance on multiple-choice is itself an equity problem, as is the use of high-stakes exams.

Standard 3: Professional development.

The state does not require assessment education for pre-service or in-service teachers. Training is available to educators in a full range of performance and classroom assessment practices and the state exams. Evaluating teacher competence in assessment is one component in the state model for local evaluation of teachers. The SEA has not surveyed educators to determine if their assessment professional development needs have been met.

Evaluation. It is difficult to determine the actual extent of the trainings in classroom assessment; they should be extensively available if they are not. A survey could inform the SEA about unmet needs and interests. Pre-service requirements in classroom assessment should be part of teacher education. Making competence in assessment part of the evaluation of teachers is a good idea; the state should study how well such evaluation is done and whether it has a positive impact.

Standard 4: Public education, reporting and parents' rights.

Test scores can be appealed and items challenged on the Achievement Test and Competency Tests (not writing), but the assessments cannot be reviewed by parents, nor can parents opt their children out of testing.

Results of the Achievement Test are reported to schools and parents in 3-4 weeks, to the public in 4-5 months. Results for writing are reported in 8 weeks to all stakeholders, for Competency Tests in 6 weeks. Guidance on using results is provided to all stakeholders. Explanatory information is provided to parents and the community on the writing assessment. For the Achievement Test, but not for the other tests, the SEA has surveyed parents and the public to determine what information they want reported and whether the reports are understood.

Reporting includes the Tennessee Value-Added Assessment System (TVAAS), which reports results in terms of school and district progress compared to their past performance, using three-year rolling averages. The goal is "for Tennessee student gains to equal or exceed national norm gains in each subject by the end of this century."

Evaluation. The right of parents to review tests and exempt their children from testing should be built into the system. Guidance on use and explanatory information seems adequate given the limited nature of the tests. The TVAAS is complicated and the state acted appropriately to survey parents to determine whether reports are understood. As TVAAS will be used with the new high school subject matter tests, the state needs to continue investigating whether reports are understood by the public.

Standard 5: System review and improvement.

The state has not evaluated district, school or classroom assessment practices. It has reviewed the TVASS once, using external reviewers. The review did not consider the impact of assessment on curriculum, instruction or high school graduation of the various tests, nor was the ability of the tests to assess critical thinking evaluated.

The Achievement and Competency Tests are not intended to guide curriculum or instruction. The state reported that first-time passing scores have been increasing. Annual surveys of over half the students in grades 2-5 and 6-8 are attached to the Achievement Test, providing a basis for correlating scores with various factors. The survey found that multiple-choice tests were the most common form of classroom assessment in grades 6-8, and their share of assessment is growing.

No effort has been made to take account of different learning styles. Tests for young children have not been reviewed for developmental appropriateness. None of the tests have been independently reviewed for alignment to standards, but achievement and writing have been reviewed by the contractor. The writing test is intended to improve instruction. While no studies have been done on instructional practice, the writing scores have improved.

Evaluation. Reviews have not been adequate and need to be substantially strengthened, particularly regarding the impact of exams on curriculum, instruction, and student progress. Increasing test scores is not sufficient evidence of improved learning; they may be inflated due to teaching to the test. The state testing program, which focuses on multiple-choice, is apparently mirrored in school practices, suggesting the narrowing impact of the tests on classroom practices. More investigation of this issue is essential, including a focus on how the tests affect instruction in critical thinking and cognitively complex work in the various subject areas. The tests for young children need to be evaluated for developmental appropriateness, including for their consequences for instruction.

Tennessee responded to the full FairTest survey and sent various reports. This report also used CCSSO/NCREL, CCSSO and AFT reports. The state replied to a draft description.

TEXAS

Summary evaluation.

The Texas assessment system needs many major changes. It relies almost entirely on multiple-choice items, except for a writing prompt, and has a high-stakes graduation test. On most of the other standards, however, the state does very well. It has strong bias review procedures, provides solid public information, accords parents substantial rights, and has a thorough and continuing review system. Professional development appears fairly extensive.

Standard 1: Assessment supports important student learning.

State content and performance standards are being revised or developed in English language arts/reading, math, science, social studies, fine arts, health and physical education, languages other than English, technology applications, and other areas. Educators, parents and other stakeholders were involved in drafting the standards. The SEA reported that the assessment frameworks have been completed.

The Texas Assessment of Academic Skills (TAAS) is a CRT that assesses reading and math in grades 3 through 8, writing in grades 4 and 8, and science and social studies in grade 8. TAAS reading, writing and math tests are also used as a high school graduation requirement, with administration beginning in grade 10. End of course exams exist for Algebra I and Biology I, and others are in development. Other than the writing assessment, the exams are all machine-scorable, mostly multiple-choice. New end-of-course tests in English II and US History will include constructed-response items. Writing assessments require responses to commercially-developed, SEA-approved prompts. Papers not meeting the minimum standard receive more detailed scoring to provide feedback to the students. All TAAS tests are untimed.

Texas now has Spanish versions of the TAAS in reading and math in grades 3-6 and in writing in grade 4. The SEA also recently developed a reading inventory to be used in the classroom at local discretion at grades K-2. The results will not be reported to the state. The state also provides for voluntary assessment of students in private schools and home schools.

Results of assessments are used for student diagnosis or placement, instruction and curriculum improvement and program evaluation. Students, with some exceptions, must pass the TAAS high school exit exam to receive a diploma. The state contracted for development of a proposal for an alternative assessment system for students who did not pass the TAAS high school exam. However, the proposal concluded that other than individually administering the TAAS exams (which is currently available to all students), an alternative system probably would not make a significant difference in pass rates. Districts can use TAAS tests as grade promotion gates if they choose. Consequences for schools include exemption from regulations, monetary rewards, warnings, probation, funding/accreditation loss, takeover and dissolution.

Evaluation. The heavy reliance on multiple-choice and use of a high-stakes test for graduation are the serious flaws in this assessment system. The state should shift toward performance assessments and drop the high school exit exam. The Spanish-language exams are a very positive development.

Standard 2: Assessments are fair.

Print materials of various kinds are provided to students to prepare them for the exams.

Two levels of bias review committees exist, a general content committee and a specific bias committee that includes only members of minority groups. The committees have authority to delete or modify items. Statistical analyses are done pre- and post-test.

Disaggregated data are reported at school, district and state levels by race, gender, free or reduced lunch eligibility, LEP and IEP status. IEP students' scores are only reported separately.

Seven percent of students tested are classified IEP or LEP. Extensive accommodations are available for IEP students. Procedures for alternate assessments for IEP students are under development. IEP students may be exempt from TAAS by their IEP committees, including the exit exam. If exempted, they can earn a regular diploma. Some LEP students also are exempt, but they must pass the English language arts and other high school exit tests in order to obtain a diploma. A modest range of accommodations are available for LEP students, and tests are available in Spanish in some subject areas.

Evaluation. The bias review procedures, reporting of disaggregated data and provision of accommodations are all solid. IEP students should be included in the full reports. The development of alternatives for IEP and LEP students will be positive. The use of a high-stakes test and over-reliance of multiple-choice items are not in line with this standard.

Standard 3: Professional development.

The state does have general requirements for education in assessment for colleges of education, but the colleges vary in implementation. The assessment division of the SEA does not survey teachers as to their professional development needs in assessment, but that may be done by other divisions. The assessment division does provide a variety of materials, including technical digests, test interpretation guides, videos, and a television series on assessment. Regional service centers meet monthly on professional development issues. The state has a voluntary teacher evaluation, including knowledge in assessment. LEA's can develop their own evaluations, which are not regulated by the state and may or may not include knowledge in assessment.

Evaluation. As the assessment division has only limited information about professional development in assessment, it is possible that more is occurring at both pre- and in-service levels. The efforts of the assessment division and the regional service centers seems solid. The SEA should determine teacher capability in classroom assessment and take any needed steps to improve their skills.

Standard 4: Public education, reporting and parents' rights.

Print materials are available, some in Spanish as well as English, to parents and policymakers. The state has surveyed parents and conducted focus groups around the state with parents and students to determine what information parents want reported and whether the reports are understood. Parents may appeal scores and opt their children out of testing. If a child is exempted from the high school exit test, the parents must acknowledge this means the child will not receive a diploma. Parents may examine tests after administration.

Evaluation. Texas is quite strong on this standard.

Standard 5: System review and improvement.

Texas regularly reviews its assessment program, including internal audits and the use of external experts. A full evaluation is done every two-to-three years. Regional groups of schools participate in statewide meetings every few months to provide guidance to the state assessment program. Teacher educators who help develop the exams provide another form of ongoing review.

Alignment with standards is obtained by beginning test development from the curriculum, writing items to match aspects of the curriculum determined to be most important, then reviewing the items to make sure they measure what was intended. Even prior to field-testing, Items are sometimes tried out in small pilots that include interviews with students and teachers. Teachers from varied ethnic and geographical backgrounds, and who teach a variety of different types of students (e.g., special education, gifted and talented), are involved to help ensure the tests are appropriate for students with different learning needs and different cultural backgrounds. Reviews consider developmental appropriateness for the grade 3 test.

The state says that studying the impact of assessment on curriculum and instruction is the centerpiece of its evaluations. The exams are evaluated for their ability to assess cognitively complex material. The SEA recognizes that not all aspects of the standards can be assessed with paper and pencil measures. Technical reports are done.

The reviews consider positive and negative intended and unintended consequences. On the positive side, the grade 4 writing requirements are now more advanced than were the high school requirements when writing was first assessed in 1980. On the negative side, there are teachers and schools which misuse the tests, treating them as the focus of curriculum and instruction rather than as measures that can provide some guidance for improvement.

The state has done some evaluation of the impact of the exit exams on high school graduation rates. They conclude that of those who do not graduate, about half are affected by the exam requirement (though many of these students might not graduate anyway). The data are inexact because some districts do not have good records on their students.

Evaluation. Texas extensively reviews its assessments system and its impact on education. State officials are aware of problems, such as some teachers teaching too narrowly to the test. The state has concluded that despite the problems, the overall impact is positive and that the heavy reliance on multiple-choice is a reasonable measurement procedure, a conclusion that is not shared by those who endorsed the Forum's *Principles*.

Texas responded to the short FairTest survey by telephone. This report also relied on CCSSO/NCREL, CCSSO and AFT reports. The state replied to a draft description.

UTAH

Summary evaluation.

Utah's student assessment system needs a complete overhaul. It tests far too often, relies too heavily on multiple-choice and uses NRTs. Equity concerns seem to be inadequately addressed, and professional development is insufficient. System review needs improvement. It may be that the SEA's new efforts in performance and portfolio assessment can be the basis for redesigning the system. Positively, the stakes are moderate.

Standard 1: Assessment supports important student learning.

Utah has curriculum frameworks, including standards and achievement indicators, in English language arts, math, science, social studies, arts, information technology, and health and physical education. They were developed on a collaborative basis with state and district personnel, including teachers and parents.

Utah administers the Stanford 9 NRT in grades 5, 8 and 11 in the areas of language arts, math, reading, science, and social studies. The Core Curriculum Testing program in grades 1-12 is required of districts by the state. While districts may choose instruments they will use, all use the state's Core Curriculum Assessment Program (CCAP), which includes criterion-referenced multiple-choice tests for grades 1 - 6 in reading, math and science; and end-of-course tests in science and math in grades 7 - 12. Under recently-passed legislation, starting in the fall of 1997, entering kindergartners will be assessed on early reading and counting skills.

The SEA has developed and makes available to districts a variety of individual and group performance assessments and portfolio materials in math, reading, science, social studies, and visual arts. The SEA also has a writing assessment which obtains writing samples to SEA-provided prompts. It is voluntary for both districts and students. Scoring is analytic using a model developed by the Northwest Lab and done by whomever each participating district designates.

The CCAP tests and the performance assessments were developed to assess the core curriculum. Further revision and development of assessments is planned.

Results of assessments are used for student diagnosis or placement, curriculum improvement, program evaluation, and, in the case of the NRT, school performance reporting and student awards. The SEA established "NAEP linked proficiency levels" for reporting school and student achievement. Teachers may use the end-of-course tests as final exams or otherwise as part of their students' final grades.

Evaluation. Utah relies too heavily on NRTs and much too heavily on multiple-choice. Testing young children with multiple-choice instruments exacerbates the problem. The NRTs should be dropped, particularly since the state uses NAEP as a basis for reporting student achievement in light of national standards. The state should shift toward constructed-response, performance and portfolio assessments based on the standards and away from multiple-choice testing. The testing burden should be substantially reduced. Positively, the stakes are not high, though the SEA should review how heavily districts rely on results from one test to place students or to determine high school grades.

Standard 2: Assessments are fair.

The state does not have a bias review committee or procedure for the CCAP or performance assessments. Data reporting does not include demographic categories.

Students with IEPs can be excluded from the NRT's. Decisions on the CCAP are made locally, but the tendency is to test a student if meaningful results can be obtained. LEP students can be excluded from the statewide testing if they have been taught in English for less than three years and they cannot participate meaningfully. Results for IEP and LEP students who are tested are included in regular state NRT reports.

Evaluation. Bias reduction efforts are weak and reporting should include demographic data. Testing many students with the NRT may be inappropriate. More data on actual LEA practices is needed, but the reported tendency is positive.

Standard 3: Professional development.

Professional development is recognized as a major challenge. Pre-service teachers have no specific course requirements. The state has not surveyed teachers' competence in assessment or their professional development needs. Print and video materials are provided to educators and parents for professional development purposes. Training materials for scoring the writing assessment were provided to districts.

Evaluation. Professional development needs to be substantially strengthened for pre-service and in-service teachers. The SEA should encourage districts to have the writing samples scored by teachers. We did not receive information regarding who scores performance assessments. Teachers also should be involved in the development of performance assessment items. Both scoring and writing can be avenues for enhancing professional development.

Standard 4: Public education, reporting and parents' rights.

Test item pools are made available for teachers' classroom use, but practice tests or materials are not distributed on the state-made tests. Practice tests are used in grades 5 and 8 with the NRT program.

SEA reports results on the NRT and each of its subtests by state, district, and school, and includes trend data. The reports are in English only. Districts report data on the CCAP. We do not have data on parent access to exams.

Evaluation. The adequacy of reporting should be investigated. Public education on performance assessments is likely to be needed.

Standard 5: System review and improvement.

The SEA has technical manuals on the CCAP, including reliability data and information regarding the relationship between the tests and the core curriculum. The SEA reports that it pays "close attention" to testing higher order thinking and that many of its multiple-choice items do that, but it recognizes that not all aspects can be tested with multiple-choice items. The performance, portfolio and writing assessments are in part intended to address this gap. The tests were examined for developmental appropriateness of items for younger students. The impact of assessments on curriculum and assessment has been investigated for some pilot projects, but not on a large scale. We did not receive information on the involvement of the public or of outside experts in the review process.

Evaluation. The SEA may overestimate how well the multiple-choice tests evaluate higher order thinking and the developmental appropriateness of testing young children (a review of items does not constitute a review of the method of assessment). The performance, portfolio and writing assessments seem to play a minor role, behind the NRTs and the CCAP. The actual impact of the three components on classroom practice should be examined, particularly since the CCAP is used in student grading. Detailed review of LEA practices is also warranted.

Utah responded by telephone to the short form of the FairTest survey. This report also used CCSSO/NCREL, CCSSO and AFT reports. The state replied to a draft description.

VERMONT

Summary evaluation.

Vermont has nearly a model system. Many of the improvements that should be made to solidify the program are already planned. The assessments are based on state standards, rely very little on multiple-choice items and include portfolios in two subjects. The assessment burden is reasonable, as are the stakes. Sampling is done in re-scoring some locally-scored portfolios in order to obtain state data, but will not be used on exams. Fairness is adequate and improving. Professional development is good and also improving. Public education and parent rights are solid. Reviews are good and are being strengthened. Improving the reliability of the portfolio assessments is an area that will need continued attention. Aside from further progress in the areas already planned, the state should consider using more performance tasks in addition to constructed-response items and using portfolios in additional subjects.

Standard 1: Assessment supports important student learning.

The Vermont Frameworks of Standards and Learning Opportunities are organized in two groups. "Vital results" standards cross subject boundaries and includes such things as problem-solving and habits of mind. "Fields of knowledge" has sets of standards in three areas: Arts, Language and Literature; History and Social Science; and Science, Math and Technology. Opportunity to learn standards are part of the Frameworks.

The state's standards were developed by teachers, professors in the field, business people, parents and students, with wide public participation in forums around the state and with involvement of teachers in each discipline. They are intended to guide local standards and curriculum development. Recently, a state Supreme Court decision which found the state's educational funding unconstitutional has led the legislature to rethink the state's education programs, including the provision of equal opportunity to learn and the use of standards.

Vermont's assessments are voluntary for districts through this year, but next year will become mandatory for public schools and for private schools that receive state funding. The state assessment program includes locally-scored portfolios in grades 4, 8 and 10 in math, and 5 and 8 in writing. To obtain state data, the state re-scores samples from each school in each subject in alternating years. Districts score a sample in each subject annually to obtain district and school scores. Students have the final say about what goes into their portfolios, but use the state content requirements, the scoring guides, and teacher advice in their selections. Half the schools in a statewide survey reported using the portfolios across the grades, not just in the grades in which the state collects data.

The state is introducing New Standards Reference Exams (NS) in math and English language arts (ELA) in grades 4, 8 and 10. The SEA reports that the NS exams are aligned with the state's standards because Vermont participated in developing NS standards, and the NS and state standards were developed together. The NS math exam contains a mix of short to longer constructed-response items, while the ELA exam is about one-third multiple-choice and the rest constructed-response.

With a contractor, the state is developing a science exam, for grades 6 and 11. It was piloted in 1996 and will be introduced in spring 1998. It will be mostly constructed-response, with some multiple-choice. The state plans to develop an assessment in history and social science, for grades 6, 9 and 11, that also will be mostly constructed-response.

The state used matrix sampling for the science exam pilot, but because nearly half the state's elementary schools have fewer than 20 students in a grade, statistically significant data could not be obtained for many schools. Thus, the state will not use sampling.

Results of assessments are used for curriculum improvement, program evaluation, and school performance reporting. Schools release their own reports based on the portfolio results, and the state releases a state report. Under pending legislation, each school will be required to develop action plans for improvement that make use of student assessment results. Under federal Title I legislation, assessment data will have to be used as part of school progress evaluations.

Evaluation. In using portfolios and exams that are mostly constructed-response, the state is close to a model system. The extension of portfolios to other subjects or the addition of performance tasks would make the system even better. Keeping the stakes moderately low while using assessment results for improvement is reasonable. The assessment burden is fairly light and spread across a number of grades. The recommendation in these Standards for sampling does not appear to be technically feasible due to the many very small schools.

Standard 2: Assessments are fair.

Vermont will be creating a bias review committee. The state has very few racial or linguistic minorities. Thus far, data have been examined for gender bias. The intent is to look at SES also, but the means of obtaining SES data are not yet established. New Standards is responsible for reviewing its exams.

IEP students may be excluded from the assessment. The state has provided limited guidance on this process thus far, but will provide more in the future. Accommodations are available for LEP, and under the new IDEA law alternatives to the NS exams will have to be developed. The intent is to appropriately assess all students. Those assessed are included in regular reports.

For NS exams, sample tasks are released for practice. The state also views participation in portfolios as good preparation for the NS exams. Portfolios are integrated into classroom work, and the scoring guides are available.

Evaluation. Once the state has created its bias review committee, extended bias review to consideration of SES, and ensured full inclusion in assessments, the state will meet this standard.

Standard 3: Professional development.

The state does have pre-service assessment requirements which are being changed to better match the requirements of portfolios and performance assessments. New legislation supports the changes. Vermont has employed extensive professional development around its use of portfolios. Currently, the state has networks, a small group of teachers in each portfolio subject who develop materials on using portfolios that are widely used by teachers. A Science Initiative funded by the National Science Foundation includes professional development in assessment. Some evidence suggests that the portfolios have also had a positive effect in other subjects, as well as in grades that are not part of the state scoring.

The SEA is now attempting to implement more systematic professional development. An initial planning meeting to develop coordination in professional development was held in the spring of 1997 and work will continue. The professional development changes are part of new legislation that will require teachers to be able to teach to the standards for licensing and re-licensing. The SEA expects the regulations will include assessment capability, including both state and classroom assessments.

Evaluation. The state will be strengthening what was already one of the nation's more extensive professional development efforts. The portfolios were intended to improve teaching as well as to assess student learning, and independent reviews, such as by the RAND Corporation, have confirmed that this intent has been met.

Standard 4: Public education, reporting and parents' rights.

The state makes available standards and examples of student work. Districts and schools are responsible for annual reporting, and the state will be providing more guidance on this. A web page includes data on every school in the state.

Parents can review the NS exams at the schools only (items are reused). No formal challenge mechanism exists, but scores can be reviewed if parents raise questions. Parents have been able to opt their children out of state testing.

Evaluation. Public education about Vermont's system has been fairly extensive and solid. Parental and student rights are reasonable. The state's stronger guidance in district reporting should lead to further improvements.

Standard 5: System review and improvement.

Vermont has not had a systematic review process, but is planning to develop one. It has done some relevant studies, and others have been conducted by outside evaluators. The planned evaluation system will focus on results for students and on support for standards, including assessment. It will include forums and site visits to study implementation. The assessments' impact on changing the curriculum will be the focus of future study. Technical studies have been done, and more are planned.

The SEA reports that some areas of state standards cannot be measured with paper and pencil assessments. Alignment of NS exams and state standards was ensured through Vermont's participation in the NS standards development, and a linking study was part of that process. The portfolio rubrics were created before the standards, but the standards incorporate references to the portfolio rubrics. The writing portfolio content requirements may be revised, based on reviews and the state standards.

One nearly-completed study has found that students in schools that used math portfolios across the grades performed better on problem solving (the focus of the portfolio) on the NS math exam than did students in schools that only use portfolios in the state-scoring grades or did not use them at all. One study of task difficulty was done in math and will be done again next year.

Evaluation. Reviews have existed but have not been systematic, so the plan for more systematic evaluation is important. It is also important that the state ensure that all the assessments, not just math, are measuring cognitively complex material and critical thinking skills, as they are intended to do. In the past, technical studies have found problems with the reliability of Vermont's portfolio program. While improvements have been made, the SEA will need to continue to focus attention on strengthening reliability. Changes in the writing portfolio that are being considered may help in this regard.

Vermont responded by telephone to the short form of the FairTest survey and to some additional questions. This report also used CCSSO/NCREL, CCSSO and AFT reports. The state replied to a draft description.

VIRGINIA

Summary evaluation.

Virginia's system needs a complete overhaul, but the revisions the state is making are not positive. The state tests too often and will soon test more, relies entirely on multiple-choice except for writing to prompts, mandates use of an NRT, and has a high school graduation test requirement. The state needs to drop these and its new multiple-choice tests, and develop an assessment system for a few grades using a mix of methods. Bias reduction is borderline adequate and inclusion of IEP and LEP students is being worked on. Professional development is thoroughly inadequate. Public education is acceptable for the kinds of tests used, and reporting is adequate, with surveys of parents and the public a positive point. Reviews are extensive, but apparently fail to address basic issues surrounding the limitations of relying on multiple-choice tests.

Standard 1: Assessment supports important student learning.

Virginia has content standards and student expectations developed by local educators, the state board of education and other stakeholders, in math, science, English, history, and technology. The state is in the process of developing assessments to measure its content standards.

The Virginia State Assessment Program (VSAP) currently includes the abbreviated Stanford 9, a multiple-choice NRT, given to all students in grades 3, 5, 8, and 11 in the areas of math, reading and language. Virginia also administers the Literacy Passport Test (LPT), which consists of criterion-referenced tests based on 1988 standards. In the LPT, reading comprehension is tested by a customized Degrees of Reading Power test (multiple choice, using the cloze method). Math is tested with a state-developed multiple-choice test. The writing assessment uses SEA provided prompts (one per sitting, twice a year, with an optional third administration in the summer) in which students are asked to produce a writing sample on an untimed basis ("but within the course of a single sitting"). They are scored by a commercial company in five domains. All students are tested and multiple forms are used. The LPT is first administered in grade 6 and is given in subsequent grades to those who have not passed it and to transfer students. Students must pass all three components to earn a standard diploma.

The state is currently field-testing a new Standards of Learning (SOL) assessment, developed by a contractor based on the 1995 standards. The SOL will be administered in grades 3, 5, 8 and 11 in English, math, history and science, as well as technology in grades 5 and 8. The tests will be multiple-choice with writing to a prompt, similar to the current LPT, in grades 5, 8 and 11. The tests may become part of the state's graduation requirements.

Results of the LPT are used for student diagnosis as well as graduation. Students who do not pass the LPT by the end of eighth grade cannot be classified as high school students (though they may take high school courses) and are not allowed to participate in interscholastic activities (e.g., hold class office, play varsity sports, or join the debate team). All such students must have individual Learning Development Plans to guide their studies to enable them to pass the test.

Parents of home-schooled children must provide evidence of achievement, which usually means the results of a standardized test.

Evaluation. Virginia fails to meet this standard since it tests too often, relies on multiple-choice on both current and prospective tests (except for writing samples), administers an NRT, and has high stakes. The current high school exit exam is based on now-outdated standards, which means students will be tested to two different sets of standards.

Standard 2: Assessments are fair.

A bias/equity review committee which represents racial/ethnic and gender interests reviews all items on state-made tests. The committee may recommend that an item be modified or removed from the bank. Test results are disaggregated by gender, ethnicity, disability status and English-speaking status.

About 14 percent of students tested are classified with an IEP; 1 percent are LEP. All students must pass the LPT to receive a standard diploma, there are no exceptions. However, IEP students are not required to pass the LPT to be classified as high school students, and accommodations or postponing test-taking are allowed for IEP students. LEP students who have not been in a Virginia public school for three years can be classified as a high school student without passing the LPT, but must pass the test after being in school for three years. LEP and IEP students who are tested are included in regular state reports. Accommodations on the NRT are available for both LEP and IEP students, but their scores are excluded from regular reports if an accommodation which does not maintain standard conditions is used. Further accommodations for IEP and LEP students on the NRT and accommodations on the SOL are being developed.

Evaluation. The bias review committee could be strengthened. Reporting by demographic groups is positive. The extent and quality of accommodations are still being developed. Reliance on multiple-choice does not meet this standard as it does not allow for variations in learning styles or cultures. The graduation exit test also fails this standard.

Standard 3: Professional development.

Pre-service teachers are required to learn about standardized tests and use of test results. Other assessment training is available in the state for in-service teachers. The state does not evaluate teacher competence in assessment or survey to determine teacher training needs.

Evaluation. Professional development is inadequate. Pre-service requirements should be expanded to include classroom and performance assessments. Training in these areas should be systematically provided to in-service teachers.

Standard 4: Public education, reporting and parents' rights.

Descriptions of assessment methods, scoring guides, and samples or examples of work, including sample LPT items, are distributed to students, teachers, administrators, parents and the community, using print, video, TV broadcast, meetings and trainings. Students can appeal their scores. Requests to review an assessment are handled on an individual basis.

Results of the VSAP and LPT are publicly reported at the school, district and state levels. Results of the LPT are reported within 90 days after administration, the NRT in 3 to 4 months after administration, to students, parents, schools and the public. Results for both are reported only in English. Parents and the public have been surveyed to find out what assessment information they want and whether they understand the reports.

Evaluation. Public education is acceptable for the kind of assessment programs Virginia has. Reporting is acceptable, and the surveying is positive.

Standard 5: System review and improvement.

Studies of the consequences of the LPT for curriculum and the impact of the LPT on graduation have been conducted. Content validity and reliability for the LPT have been researched. The SEA has not surveyed district, school or classroom assessment practices, but has surveyed districts and schools regarding remedial programs for the LPT.

Both new state-administered assessments (the Stanford 9 and SOL field-tested items) have been reviewed by content review committees, comprised primarily of classroom teachers, instructional supervisors and university faculty for their match to Virginia's Standards of Learning, both in terms of content and intended cognitive demand.

Evaluation. Conducting studies of validity and impact on curriculum and graduation is positive. Matching the assessments to the new standards also is positive, but as the tests are multiple-choice, either the standards are not fully assessed by the tests or the standards are not adequate.

Virginia responded to the full FairTest survey and sent various documents. This report also used CCSSO/NCREL, CCSSO and AFT reports. The state replied to a descriptive draft.

WASHINGTON

Summary evaluation.

Washington's current system needs many major improvements, particularly discontinuing its use of an NRT. Many of its problems currently are being addressed. The state is developing a far more positive system that will use multiple assessment methods. Among the goals are to minimize the amount of testing, keep the stakes relatively low, emphasize development of classroom-based assessments, provide a strong commitment to professional development and include strong plans for system review. Continuing concerns include ensuring that an NRT is not used, multiple-choice items are a minor part of the exams, teachers are heavily involved in developing and scoring assessments, parents and the public are educated about and involved in the new assessments, and exams do not become single, high-stakes hurdles for the planned Certificate of Mastery.

Standard 1: Assessment supports important student learning.

Washington is developing *Essential Academic Learning Requirements*, content standards in reading, writing, communications, math, science, social studies, art, and health. The state is planning to align new assessments to these emerging standards and expects the new assessments to guide curriculum and instruction.

The state's current Basic Assessment Program includes an NRT battery, the CTBS/4, in grades 4 and 8, and the customized off-the-shelf criterion referenced Curriculum Frameworks Assessment System (CFAS), in grade 11 in math, English/language arts, science, and social studies.

The new assessment system is in the second year of a planned 5-year development program. The system will include state-level exams, with multiple-choice, short-answer and extended, constructed-response items. They will be administered in grades 4, 7 and 10. Currently the grade 4 math and communications (reading, writing, listening) state exam is operational and voluntary for districts. The grade 7 tests in the same subjects are in pilot stage, to be operational next year, and the grade 10 tests are in planning, to be piloted next year. A science assessment will soon be under development. The writing portion of the communication exam requires two pieces, one longer with revisions and one shorter. For cost reasons, scoring will be done by a contracted company, not by state teachers. Three-point rubrics for short responses and 4-point scales for extended responses are being developed.

A second part of the new system will be classroom-based evidence, which is in the planning stage and is closely related to extensive new professional development activities.

Results of current assessments are used for student diagnosis or placement, curriculum improvement, program evaluation and school performance reporting. Consequences include funding gains for schools.

Much about how the state will use the results from the new exams is being worked out. The state will require a certificate of mastery, which probably will be based in part on the grade 10 exam. The certificate likely will be necessary but not sufficient for a diploma. Exam results may be used for school accountability, but a planned new state indicator system is just beginning development. Whether the state will continue to use an NRT also will be decided later.

Evaluation. The current system relies too heavily on multiple-choice items and the NRT. The shift toward a mixed-method system is positive, though it is not clear what the proportions will be (multiple-choice should be a minor part). The state should drop the NRT, both to focus assessments on the standards and to prevent the testing load from becoming excessive. Performance tasks and the writing samples should be scored by teachers. The uses of the new assessments also remain unclear. The exam should be only a part of determining receipt of the certificate, as is now planned. The classroom-based part of the new system, linked to professional development, appears promising.

Standard 2: Assessments are fair.

Bias review, including an item-review committee and technical analysis of items, will be used in the new assessments. Flawed items will be discarded. An item data bank large enough for five years of exams will be reviewed. Gender, race and SES data already is collected and reported for schools, districts and the state.

Currently, 3 percent of students tested are classified with LEP, 5 percent have IEPs. Some students with IEPs or LEP are not tested. Limited accommodations are available for those with IEPs. Results for those tested are included in regular reports. The intent of the new system is to assess almost every student, with all accommodations used in the classroom allowed on the assessment. A statewide committee is working to develop guidelines for LEP and IEP accommodations. The state is attempting to design a reporting system with no incentives to exclude students.

Evaluation. The state has a strong approach to fairness planned for its new assessment program, ranging from the bias review committee to reporting demographic-based data to very strong inclusion efforts (though the new federal IDEA legislation will require even more). The use of multiple-methods in the assessment and the focus on developing classroom-based assessments should also bolster equity. The state consider developing assessments in languages other than English.

Standard 3: Professional development.

Washington requires testing, assessment and evaluation knowledge of all prospective teachers. The SEA has an extensive professional development program consisting of 16 regional training centers where teachers learn new assessment techniques and then train others in their districts. Training focuses on teachers' understanding of statewide assessment, mastery of standards and ability to choose and use appropriate classroom assessments. Professional development and the new assessments are seen as complementary.

Evaluation. The state has a strong and commendable commitment to professional development. Teacher involvement in scoring would strengthen this. Whether the professional development efforts will prove sufficient for the new programs remains to be seen. The state should survey teachers and administrators periodically, perhaps as part of the student assessment or as continuing surveys on teacher competence in assessment, to ascertain educator needs.

Standard 4: Public education, reporting and parents' rights.

All students are expected to be tested. However, some parents have exempted their children from state assessments. Less than 1 percent opt out, but the number is reportedly growing. We did not receive information on other aspects of this standard.

Evaluation. Too little information to evaluate.

Standard 5: System review and improvement.

The state has surveyed district assessment practices and recently surveyed teachers' attitudes toward current testing to obtain baseline data so that teacher knowledge and attitudes can be studied as the new assessment system is implemented. The law requires the SEA to study the impact of the new assessment system on curriculum and instruction. The intent is for the new assessments to be able to assess higher order thinking in subjects. A concern about too much testing in grade 4 is being addressed.

Evaluation. Most of the right questions are being asked of the new assessments, so the evaluation planning process seems on target. Public and outside expert involvement should be included, along with technical reviews and plans to use the information systematically to improve assessment. The functioning of the classroom assessments and teacher competence will require particular attention.

Washington responded by telephone to parts of the short form of the FairTest survey. This report also relied on CCSSO/NCREL, CCSSO and AFT reports. The state replied to a draft description.

WEST VIRGINIA

Summary evaluation.

The state assessment system needs a complete overhaul. The state testing burden is far too heavy and relies on multiple-choice NRTs. The number of grades tested should be drastically reduced and the state should shift away from using an NRT. The state should develop an assessment system that more fully matches the standards and that relies on a variety of methods of assessment. An NRT which only partially matches the state standards should not be the basis for mandated reteaching or the awarding of "warranties" on diplomas. Bias review and inclusion are insufficient, professional development needs to be strengthened and reporting is not sufficient. The assessment review process is inadequate, though the on-site visits are a positive approach.

Standard 1: Assessment supports important student learning.

West Virginia has new, revised standards in language arts, math, and science and social studies and is developing performance standards in language arts and math. The state plans to "fine tune" the assessments to better match the new standards.

By law, students are tested in every grade -- with a readiness test in kindergarten; an assessment in grades 1 and 2 for reporting to parents and in-school use; and an off-the-shelf, multiple-choice NRT (Stanford 9) in grades 3 - 11 using the full battery (reading, language, mathematics, science and social studies). The Stanford 9 has a 60 percent match to the curriculum frameworks. The SEA administers the ACT Explorer in grade 8 and the ACT Work Keys in grade 12, both related to career interest and academics. The state also has a criterion-referenced test it makes available to districts to use in grades 1-8.

A writing assessment in grades 4, 7 and 10 uses responses to SEA-provided prompts. Students are given 60 minutes at grades 7 and 10 to produce a writing sample on demand with revisions permitted. At grade 4, students have two 45-minute sessions. Designing prompts, developing rubrics and scoring are done by teachers and the standards committee. The SEA provides training to teachers for using the rubric. All students are tested unless exempted by IEP.

Results of the writing assessment are used for curriculum improvement and program evaluation. No individual or school consequences are attached. Accountability for schools includes school performance reporting on the NRT. School consequences from the NRT can include probation or loss of accreditation. Districts are required to do an item analysis of individual student results and develop a re-teaching plan for students who have not reached the fiftieth percentile by grade 10. Students who do not reach that level by graduation do not earn a state "warranty" on their diplomas.

Evaluation: West Virginia tests students too often, a problem that is compounded by reliance on a multiple-choice NRT that only partly assesses to the state standards. Such a test should not be the basis for mandated re-teaching or the award of warranties on diplomas. Teacher involvement in the writing assessment is positive.

Standard 2: Assessments are fair.

A bias review committee does not exist for writing assessments because, we were told, "there is not enough cultural diversity in West Virginia." Prompts are not studied for bias. Other tests are commercial and rely on the manufacturer's bias reduction procedures. Reports do not include results by demographic categories.

Nearly 10 percent of test-takers have an IEP. Some exemptions and accommodations are available for students with an IEP. Their scores are excluded from regular reports. No separate report is issued.

Evaluation. It may be that there is little racial diversity in West Virginia, but cultural variations rooted in socio-economic class should be considered. Different learning styles are not addressed due to reliance on multiple-choice items. Inclusion for IEP appears weak, and no information was provided about LEP students.

Standard 3: Professional development.

The state provides courses and in-services on a wide range of assessment issues for teachers and administrators, but has no particular requirements. It does not evaluate or survey teacher competence in assessment or teacher training needs.

Evaluation. Professional development needs to be strengthened. Teacher involvement in writing is positive.

Standard 4: Public education, reporting and parents' rights.

The state has not surveyed parents or the public to find out what information they want reported or if they understand the reports. No particular information about the writing test is provided to educators, students or the public. On the NRT, practice tests are available.

Results of the writing assessment are reported in 6 months after administration and the results of the NRT in 6 weeks to students, parents, schools and the public. Reports are in English only. The state provides guidance on the use of results to teachers, school administrators, psychologists/counselors and district administrators for both components. Students and parents can review the NRT.

Evaluation. Public participation is limited as is public information and reporting. Allowing parental review of the NRT is positive.

Standard 5: System review and improvement.

Media reports have charged the SEA with aligning the standards to the test. The state denies this, stating that the test was selected as having the best match of any off-the-shelf to the state's standards. The SEA conducts random visits to schools in one quarter of the counties each year. Among other things, it reviews materials and lesson plans to see if instruction is covering all the curriculum areas, not just those measured by the test.

The SEA has not evaluated assessment practices at district, school or classroom levels. The state system is reviewed annually by the SEA. While the NRT is intended to guide curriculum and instruction, the impact on curriculum, instruction or graduation rates is not studied, nor are consequences such as tracking or grade retention. Tests have not been studied for developmental appropriateness.

Evaluation. Review, while conducted annually, does not address fundamental issues such as impact on curriculum and instruction. The SEA does know that at least 40 percent of the standards are not assessed. The standards not assessed probably include critical and complex thinking in subject areas. The developmental appropriateness of testing young children has been raised as an issue in West Virginia, which decided not to use the Stanford 9 in the early grades. The visits to schools can serve an important accountability function, but it would be preferable if a more systematic study of the impact of testing on curriculum and instruction were conducted.

West Virginia responded to the full FairTest survey. This report also used CCSSO/NCREL, CCSSO and AFT reports. The state replied to a draft description.

WISCONSIN

Summary evaluation.

Wisconsin's assessment program needs many major improvements. The state relies primarily on a 60-percent multiple-choice NRT that is not aligned with the standards. The state should shift to a primarily performance assessment based on the state new content standards. The state does not have a high testing burden and the stakes are moderate. Professional development needs major attention. Some fairness concerns are well addressed, others, particularly inclusion of IEP and LEP students, less so. Public reporting and system review appear to be adequate.

Standard 1: Assessment supports important student learning.

Wisconsin is in the process of developing content standards. It has curriculum guides that are also used to help districts develop standards.

The Wisconsin Student Assessment System (WSAS) has two major programs: Knowledge and Concepts Examinations (KCE) and the Reading Comprehension Test (RCT). The state has authorized development of performance assessments, but currently they are not funded. The SEA has produced a manual to help districts develop performance assessments. It also participates in various CCSSO assessment consortia. The SEA acknowledges that some aspects of the curriculum are not assessed.

The KCE is a commercial NRT, currently the CTBS 4 ("Terra Nova"), which combines multiple-choice (approximately 60 percent of score) with some short constructed-response items (30 percent) and a writing sample (10 percent), administered in grades 4, 8 and 10. Each student has about 45 minutes to respond to a writing prompt selected by the SEA from options provided by the publishing company, which also scores the responses. Proficiency levels on the exams are currently being set to comply with federal Title I requirements. The CTBS 4 was selected for having the best match with state standards.

The criterion-referenced RCT is a multiple-choice test to which some short constructed-response items (10 percent) have been added on a pilot basis. It is developed annually with participation from teachers, administrators, SEA staff, parents, education organizations and community groups. The RCT is reviewed to ensure it is aligned with the state curriculum guides and standards. Scores on the RCT are reported in relation to other factors that affect reading comprehension, such as previous knowledge and reading strategies.

The purposes of the WSAS are to provide expectations for students and obtain data based on the expectations, promote high-quality curriculum and instruction, assist educational planning for students and identify low-performing schools. The RCT also is intended specifically to allow early evaluation of the effectiveness of school reading programs. The assessments are used for identification of pupils needing remediation, public reporting, and to prompt the development of remediation plans for any district in which fewer than 80 percent of the students score above the performance standard on the RCT.

Evaluation. Positively, the state testing burden is not high, stakes are moderate, and the NRT includes some constructed-response items. However, an NRT is an inadequate method of assessing student attainment of standards. Multiple-choice and short-answer items are inadequate for assessing some aspects of the standards. It is unfortunate that the performance assessments have not been developed. The state should be using a standards-based, predominantly constructed-response assessment rather than the CTBS. The RCT is apparently respected by educators; it is a positive sign that this test is beginning to include constructed-response items.

Standard 2: Assessments are fair.

Bias analysis is conducted on the KCE by the contractor. For the RCT, a bias review committee composed of representatives of minority groups in the state participates in selecting reading passages and items. It has the authority to delete or revise items. Scores on WSAS and RCT tests are released by race and gender at the state, district and school levels. RCT test scores are also disaggregated by the percentage of families in the district receiving AFDC and by district size.

IEP and LEP students may be excluded or may receive accommodations to take the exams. About 10 percent of third graders are excluded from the RCT. On the KCE, scores of IEP and LEP students are not reported with other students. Approximately one-half (in grade 4) and one-quarter (in grade 8) IEP and LEP students do not take the test, and those exempted comprise 3-8 percent of the student population.

Evaluation. Bias review and reporting of data are solid. The exclusion of too many students should be addressed by providing further accommodations or alternative assessments. Including more methods of assessment could enhance fairness.

Standard 3: Professional development.

The state provides professional development in assessment to teachers and administrators about the KCE and the RCT through informational workshops for district staff. The state has not attempted to evaluate district school or classroom assessment practices. Educators are involved in developing the RCT assessments, but not in scoring assessments.

Evaluation. Professional development appears inadequate for pre- and in-service teachers. The only positive is teacher involvement in developing the RCT.

Standard 4: Public education, reporting and parents' rights.

Information about assessment methods and sample assessments, as well as guides about the assessments, are provided to students and parents. Teachers and administrators also receive scoring guides and examples of student work. For the RCT, a handbook and brochure describing the RCT are provided to educators and parents, and previous years' tests are available.

On the KCE, scores can be appealed, students may challenge test items and tests may be reviewed after administration. A parental waiver provision exists on the KCE.

Assessment results and reports are available only in English. Parents and the public have not been surveyed about assessment information quality or needs.

Evaluation. The state provides substantial information to stakeholders and accessibility for parents. The state should survey the public to determine if the information is sufficient and understood.

Standard 5: System review and improvement.

The state assessment is reviewed every three years. The RCT is reviewed annually by an advisory committee which includes local educators. The tests' impact on curriculum and instruction is part of the review. On the RCT, the reading passages are reviewed for developmental appropriateness.

Evaluation. The review seems adequate. The developmental appropriateness of the grade 3 test, which is mostly multiple-choice, and the grade 4 test, which is a long exam for students of that age, should be considered more thoroughly. Item review is not the same as review of the format of the assessment. As the new state standards come into effect, the alignment between standards and the exams, and the impact of the exams on instruction toward the standards, will need to be carefully studied.

Wisconsin responded to the full FairTest survey and sent various documents. This report also relied on CCSSO/NCREL, CCSSO and AFT reports. The state replied to a draft description.

WYOMING

Summary evaluation.

Wyoming currently does not have a state assessment system that can be evaluated, It only has a vocational test administered to a sample of students. Beginning in 1998, districts will be required to assess students on the state's common core of knowledge and skills. The approach to the vocational exam seems reasonable. A district-based approach is reasonable, but the state should then evaluate the quality of the district assessments. No equity data are available, but should be. Professional development should in any event be strengthened. Public reporting is adequate for the vocational test, and the review process seems reasonable.

Legislation has just passed to create a state exam program that would assess in grades 4, 8 and 11. The SEA and an appointed committee are to report on what the SEA intends to do and the cost. In the same legislation, the SEA is ordered to study whether the state should have a high-stakes "competency test," and the logistics of implementing it. The state should utilize the *Principles and Indicators for Student Assessment Systems* to help guide the creation of the new program. A high-stakes "competency test" should be strongly resisted.

Standard 1: Assessment supports important student learning.

Wyoming is developing "student expectations" in reading, math, writing, science and social studies. They are being developed by local and state study groups with assistance from the Mid-Continent Regional Education Laboratory. State accreditation standards require school districts to develop standards and assessments by the 1997 school year in the common core of knowledge and skills.

Wyoming currently does not have a state assessment program, but it does have a vocational assessment. Legislation is pending to adopt a state assessment for all students in grades 4, 8 and 11 in reading and language arts. It will be standards-based and include multiple-choice and constructed-response items and some performance tasks.

The vocational assessment is an applied-performance assessment based on state standards. It is administered as a stratified random sample, given in grades 9-12 to about 50 percent of the students. There is no individual feedback to students. The information is used for planning and program improvement, and state studies show it affects curriculum positively. Rubrics for the assessment are developed with participation from teachers, administrators, SEA staff, outside experts and business groups. Scoring is done by teachers and SEA staff. Rubrics are made available to all vocational teachers before the assessment. Factor analysis is used to evaluate what is assessed, including critical thinking, and technical studies are conducted on the test.

Evaluation. The vocational assessment appears reasonable. As there is no other state assessment, the basic questions are what do districts do and what will the new assessments be like?

Standard 2: Assessments are fair.

There is no bias review of the vocational assessment.

Evaluation. Bias review should be done, and data on LEP and IEP students should be compiled for the vocational assessment. These issues should be addressed in a new state assessment system.

Standard 3: Professional development.

The state requires no assessment expertise for pre-service teachers, nor does it evaluate teacher competence. It has a mandatory in-service on the vocational assessment for vocational teachers that is also offered to administrators. Education in various forms of assessment are available in the state, but the SEA itself only offers sessions on the use of test results.

An annual survey about education needs includes questions about professional development in assessment. The state reports high interest in help with assessment. The state does not survey classroom or district practices.

Evaluation. Professional development should be substantially strengthened, with or without a new state exam. The survey should be a helpful starting point. Teacher involvement in the vocational assessment is positive.

Standard 4: Public education, reporting and parents' rights.

Results from the vocational assessment are reported in three months, with guidance in the use of results provided to teachers and administrators.

Evaluation. Outside involvement in developing the assessment is positive. Since the use of the test is for program evaluation, the reporting seems reasonable. If the new assessment uses performance tasks, public education about it will be necessary.

Standard 5: System review and improvement.

The state reviews assessment at all levels, including the impact of assessment on curriculum and the graduation rate. Administrators, SEA staff, outside experts and business groups participate in the review of the vocational assessment.

Evaluation. The review processes seem reasonable given the absence of a state exam. The impact of new assessments is likely to be pronounced and should be carefully and regularly reviewed.

Wyoming responded to the full FairTest survey. This report also used CCSSO/NCREL, CCSSO and AFT reports. The state replied to a draft description.

APPENDICES

Appendix A

Abbreviations

ACT -- formerly American College Testing, which makes college entrance exams and some K-12 assesssments for use by schools, such as the Explorer and Work Keys.

AFT -- American Federation of Teachers (see bibliography for their report on standards).

CAT -- California Achievement Test, a commercial, norm-referenced, multiple-choice achievement test; numbers refer to the test's edition.

CCSSO -- Council of Chief State School Officers (see bibliography for their report on standards).

CCSSO/NCREL -- Council of Chief State School Officers/North Central Regional Educational Laboratory; refers to a survey by these two groups (see bibliography for survey); unless otherwise stated, this report used the survey of the 1994-95 school year.

CRT -- Criterion-Referenced Test (see glossary).

CTBS -- Comprehensive Test of Basic Skills, a commercial, norm-referenced, multiple-choice achievement test; numbers refer to the test's editions; some editions (4 and 5) have an optional constructed-response section, but unless noted states use only the multiple-choice sections.

ESL -- English as a Second Language (see glossary).

GWU -- George Washington University Center for Equity and Excellence in Education, which has studied the participation of LEP students in state assessment programs (see bibliography).

IDEA -- Individuals with Disabilities Education Act, a federal law reauthorized in 1997.

IEP -- Individual Education Plan (see glossary).

ITBS -- Iowa Test of Basic Skills, a commercial, norm-referenced, multiple-choice achievement test for grades K-9; numbers refer to the test's edition.

LEA -- Local Education Authority, or local school district.

LEP -- Limited English Proficient/Proficiency (see glossary).

MAT -- Metropolitan Achievement Test, a commercial, norm-referenced, multiple-choice achievement test; numbers refer to the test's editions.

NAEP -- National Assessment of Educational Progress, a federal test administered to a sample of students to obtain national data, and state data for states choosing to participate.

NCATE -- National Council for Accreditation of Teacher Education.

NCREL -- North Central Regional Education Laboratory.

New Standards -- a project working with states and districts to develop standards and assessments, including mostly constructed-response "reference exams" and portfolios.

NRT -- Norm-Referenced Test (see bibliography).

SCASS -- State Collaborative on Assessment and Student Standards.

SAT -- Stanford Achievement Test, a commercial, norm-referenced, multiple-choice achievement test; numbers refer to editions; the Stanford 9 edition has an optional constructed-response section, but unless stated otherwise a state uses only the multiple-choice portion.

SEA -- State Education Authority, or state department of education or instruction.

SES -- Socio-Economic Status, classification based on level of family income, wealth, parental occupation, or related indicator.

Stanford -- Stanford Achievement Test (see SAT).

TAP -- Tests of Achievement and Proficiency, by the same publisher which makes the ITBS and which extends that test to grades 9-12.

TCS -- Test of Cognitive Skills, a NRT that purports to measure "skills important to success in the school setting."

Appendix B

Glossary

Accommodations -- changes in test administration allowed in response to needs of students with IEP or LEP, such as braille, extended time, and reading questions aloud.

Accountability -- providing to the public information about what students have learned; holding students, schools or districts responsible, such as for performance on tests; usually with specified consequences, such as denial of a diploma or rewards or sanctions for a school or district.

Adaptations -- alterations in the assessment to meet the needs of students with LEP or IEP, such as translations or using a different assessment format (e.g., a portfolio instead of a test).

Alignment/aligned -- ensuring a match between standards or curriculum and the assessment so that all items on the assessment are part of the standards and all important aspects of the standards are included in the assessment.

Alternate assessment -- used here to denote assessments designed for use with special populations for whom the regular test, even with accommodations, is inappropriate.

Alternative assessment -- any assessment that is not multiple-choice.

Bias -- a lack of objectivity, fairness or impartiality on the part of the assessor or evaluator, in the assessment instrument or procedures, or in the interpretation and evaluation process, that leads to systematic misinterpretation of student performance or knowledge based on characteristics such as race, socio-economic class, gender, or linguistic or cultural background.

Census Testing -- testing all students in the student population.

Classroom assessments -- assessments used in the classroom for diagnosis, planning, improving group or individual instruction, and evaluating student progress (distinguished from large-scale assessments).

Cloze method -- a multiple-choice testing procedure which involves leaving a word out of a passage and asking students to select a word to fill in the blank; used by the Degrees of Reading Power test.

Cognitive complexity -- in which the assessment task calls for higher level, complex intellectual activity such as problem solving, critical thinking, synthesis and evaluation of information or knowledge, and reasoning. (See also critical thinking and higher order thinking; these terms are used here somewhat interchangeably.)

Consequential validity -- or more precisely, the consequential basis of validity, in which the impacts (consequences) of using a test are considered as part of the validation of a test.

Constructed-response -- test items in which students create, rather than select, an answer; can be anything from fill-in-the-blank to extended projects or portfolios; usually refers here to medium-to-long responses, distinguished from portfolios and extended performances.

Constructivist (and social constructivist) -- theory in psychology contending that humans learn by "constructing" and revising and developing models in their minds about the world, subject matter, etc.; social constructivists also investigate the social relations that shape learning and the construction of knowledge by individuals.

Content Standards -- standards defining the desired learning (important knowledge, skills, understandings and habits of mind) in a subject area that students should acquire and be able to demonstrate; they can be more or less specific and concrete, but generally are more specific than broad goals and less specific than curricula.

Criterion-referenced test -- a measurement of achievement of specific knowledge or skills in terms of absolute levels of mastery; performance as measured against a criterion or standard (as distinguished from norm-referenced test).

Critical thinking -- the ability to problem-solve, evaluate, synthesize and reason in a subject.

Curriculum frameworks -- akin to content standards, setting out what students should know and be able to do in each subject area, but more general than actual curriculum.

Developmentally appropriate -- practices based on what is known about how children and youth develop, learn, and manifest their learning; a practice, such as assessment, that is appropriate for the developmental level of children of a certain age or of a specific child.

English as a Second Language (ESL) -- students for whom English is not their primary language or a course of study for such students

Evaluation -- the process of interpretation and use of information to make judgments.

Generalizability -- successful performance on the assessment task(s) allows making valid inferences about achievement and indicates ability to successfully perform other tasks in the subject or domain, not just the one(s) assessed.

Gridded-in -- test item in which the test-taker writes in an answer in a specified box or grid, used most often in math; akin to fill-in-the-blank.

High-stakes testing -- making important decisions, such as grade promotion or high school graduation, based on a test; some commentators argue that high-stakes tests are any tests that significantly influence curriculum or instruction, and thus activities such as reporting school scores in newspapers, which tend to push teaching to the test, are high-stakes exams; we use the term in the narrower sense.

Higher order thinking -- see critical thinking and cognitively complex.

Large-scale assessments -- assessments administered to large numbers of students, such as state or district achievement tests, as distinguished from classroom assessments.

Learning styles -- characteristic cognitive, affective and physiological behaviors that serve as relatively stable indicators of how individual learners perceive, interact with, and respond to the learning environment.

Limited English Proficiency -- students whose have limited or no knowledge of English, which limits their ability to successfully participate in an English-only educational program.

Matrix sampling -- a form of sampling in which each student responds to only a part of the whole test.

Modifications -- changes in the administration of an assessment (e.g., extended time) to make it more appropriate for some student(s).

Multiple-choice -- test items in which a student selects a response from a list of alternatives (also known as selected response).

Multiple methods -- in which more than one method of assessment is used in one assessment or set of assessments; e.g., some multiple-choice, some short-to-medium constructed-response, and some extended constructed-response or performance tasks.

Norm-referenced test -- a test which is standardized on a group of individuals whose performance is scored in relation to the performance of other individuals (contrasted with criterion-referenced test).

Open-response -- in which a test-taker constructs a response rather than selects from a list of responses (see constructed-response)

Opportunity to learn -- giving students the means to acquire the knowledge and skills for which they are held accountable; provision of equitable and adequate learning resources.

Performance assessment -- general term for an assessment activity in which students construct responses, create products, or perform demonstrations to provide evidence of their knowledge and skills; sometimes used more specifically to refer to assessments in which the student engages in an extended activity such as a lengthy essay, an extended project, a presentation to a group, a science lab, or an artistic production, as in a "performance event."

Performance Standards -- an established level of achievement, quality of performance, or degree of proficiency, specifying what a student is expected to achieve or perform to show the student has met content standards (how good is good enough).

Portfolio -- a purposeful, systematic collection of selected student work and student selfassessments gathered over time to demonstrate progress and achievement in learning.

Portfolio assessment -- the process of reviewing and evaluating student portfolios.

Professional development -- continued learning by educators to improve their knowledge and skills.

Prompt -- the topic, question, or stimulus that a student responds to on a writing test.

Psychometrics -- the attempt to measure mental characteristics, traits, knowledge, skills, etc.

Reliability -- the degree to which an assessment measures consistently or to which assessment scores are free from errors of measurement.

Rubric -- a term often used for "scoring guide" (see below).

Sampling -- a way to infer meaningful information about an entire group by examining only a representative or randomly selected portion of the group, or through matrix sampling, or a combination of both.

Scoring Guide -- a guide (or rubric) based on specified standards used to score performance assessments. Rubrics contain a scale (e.g., 6,5,4,3,2,1, or "distinguished, proficient, apprentice, novice") and descriptions of the features/characteristics of work at each point on the scale.

Stakeholders -- those individuals who have a substantial interest in schools and student learning, who may include students, teachers, administrators, other school staff, state and district education staff, parents, advocacy organizations, community members, higher education institutions, and employers.

Standards -- Statements of what student should know (content standards) and how well they should know it (performance standards).

Test burden -- the amount of testing administered in a year or school career, including number of grades in which testing is done, number of subjects tested, and length of each test.

Validity -- the degree to which evidence and theory support or disprove the adequacy and appropriateness of inferences from test scores or other assessment results and the actions based on them.

Appendix C

Methodology

A) Sources of information.

FairTest began with the 1994-95 CCSSO/NCREL survey, published in May 1996. We matched the data available from that survey to the *Principles* and discovered that many areas of the *Principles* were not covered by that survey.

We then analyzed the *Principles* to extract indicators relevant to large-scale assessments or state-level practices. We excluded areas in which information was not likely to be available. From the remainder, we constructed a fairly long survey. We asked two state assessment directors to look over the survey. In addition to suggested clarifications, one advised us that the survey was too long. While we condensed it somewhat, we decided to attempt to gather all the information we could. We mailed the survey to all 50 states in the summer of 1996. (Washington, DC, is not included in the CCSSO/NCREL survey; we sent DC both that survey and ours, but they did not respond, so they are not included in the report.) A copy of the final survey sent to all 50 states is in Appendix D.

Responses began to come in, but a few states indicated they would not participate. In the fall of 1996, we sent a follow-up letter. In early 1997, we checked with a number of states which had not replied to determine whether they would be amenable to responding to a shortened version of the survey, and a number indicated yes. The cuts were made in areas in which we had not received much information in the surveys that had been returned or in areas we decided were of less importance. A few states answered the short form questions over the telephone, rather than respond on paper. (A copy of the short-form survey is in Appendix E.) As a result of the change in the form, because some items were left blank by states, and because some states did not respond to the survey at all, the extent of the information varies from state to state.

FairTest also relied on other sources of information. We used AFT and the CCSSO reports to summarize whether a state had standards and in what subjects. News reports in media such as *Education Week* alerted us to possible changes in state assessments that we then checked, sometimes by telephone. For each state report, we list the data sources used.

Based on completed surveys, we wrote draft descriptive summaries of each state. We sent these to states to have them checked for accuracy. In a few cases, either many significant changes in the state program had occurred since the survey was first filled out or the state suggested many changes in the description. In those cases we redrafted and sent the survey back to the state for further review. In a few cases, information on standards was added after the state had checked off the descriptive draft.

For states that did not respond to the survey, we relied solely on other sources, primarily the NCREL/CCSSO survey for 1995-96 (released in June 1997), plus the AFT and CCSSO reports on standards. As a result, significant areas are not discussed for those states.

Despite our efforts to collect data on all aspects of the *Principles* and to verify that data, we recognize a series of potential problems:

- Variability in the thoroughness of state responses.
- Some information was not rechecked with the state.
- The information received depends in part on the person sending it. Occasionally we were told that material we found in other reports had never been true. Such problems may affect this report as well, though we, like others, have attempted to confirm information.
- There are state assessments that are not included in this survey. For example, some states require particular tests to be used for entrance into and exit out of programs for LEP students, but no state reported those assessments as part of the state testing program. There also may be other mandates to districts that states do not report.

Despite these potential problems, we are very confident that the data are substantially accurate and that having additional or in some cases more recent data would not alter the national findings in any significant way and only rarely would affect a state report.

Having obtained and checked the data, we subjected it to an evaluation based on the *Principles*. The grounds for evaluation and a rubric for ratings are discussed in the first parts of the section on state findings. Thus, the evaluations are FairTest's and not those of the National Forum on Assessment, which wrote the *Principles*.

B) Implications for future surveys and studies.

While the CCSSO/NCREL survey is a valuable source of information, the FairTest report includes many important areas that have not been studied by the CCSSO/NCREL survey. Topics central to the *Principles*, such as program review and evaluation, bias reduction, and professional development, are often either not included or included in only a very cursory fashion. It also is difficult to disentangle some CCSSO/NCREL data. For example, states often included their writing samples in response to questions about whether they have non-multiple-choice items in their assessments, making it difficult to determine if they had any other form of constructed-response or performance items. FairTest hopes that future CCSSO/NCREL surveys will include questions asked in the FairTest survey, making it an even more comprehensive source of data.

A major limitation of the FairTest and other surveys is the ability to use data to evaluate the actual quality of state assessments; standards; bias reduction, equity and professional development efforts; public reporting; and reviews. This is not a limitation that can readily be solved through survey methodology. Rather, it requires a more detailed qualitative analysis of state assessment programs. There does not appear to be a truly independent and representative body to undertake that important work.

FairTest's evaluations and conclusions are based on applying findings from a range of research on assessment to the available data from the states. For example, if state A uses a high-stakes, mostly multiple-choice test, FairTest's critique is based on research about high-stakes testing and multiple-choice tests and their educational impact. It is not based on a specific study of the consequences in state A. Such studies are needed, but as the FairTest survey shows, few states conduct them.

Appendix D

FairTest State Assessment Survey

Guide to Survey

1. General questions
    - 1A. Basic state data
    - 1B. General questions about the state assessment program
2. Professional Development
3. System Review
4. State Assessment Program Components
    - 4A: Additional General Component Information
    - 4B: Test Uses and Consequences
    - 4C: Who Participates in Test Development & Scoring
    - 4D: Equity Concerns
    - 4E: Reporting
    - 4F: Assessment Component Review
5. Comment

The purpose of this survey: FairTest is conducting an analysis of state-level assessment practices across the nation, in the context of the *Principles and Indicators for Student Assessment Systems* of the National Forum on Assessment (copy enclosed). Information from this survey will be combined with data from the *State Student Assessment Programs Database* (SSAPD). The resulting report will take the form of a descriptive profile of each state. FairTest is a non-profit testing and assessment reform advocacy organization. If you have questions, contact Pamela Zappardino or Bob Schaeffer at FairTest@aol.com; phone (617) 864-4810 or fax (617) 497-2224.

State_______________________________

	State Assessment Director:	Person Completing Survey:
Name		
Title		
Address		
City, State, Zip		
Telephone		
Fax		
e-mail		

1. General Questions

Person completing this section: Name______________________________________

Phone________________ Fax ________________

1A. Basic state data

If you prefer, attach and refer to report(s) which contain this information, rather than fill in the data.

1.1 Student enrollment information for 1995-96 school year:

	# students enrolled	**# in LEP***	**# in IEP****
kindergarten			
grade 1			
grade 2			
grade 3			
grade 4			
grade 5			
grade 6			
grade 7			
grade 8			
grade 9			
grade 10			
grade 11			
grade 12			
Totals			

* LEP = Program for Limited English Proficient students

* IEP = Programs for students with special needs other than LEP or gifted & talented

BEST COPY AVAILABLE

1.2 Student demographic information for 1995-96 school year:

	# students enrolled	# LEP	# IEP
African American			
American Indian			
Asian/Pacific Islander			
Latino/Hispanic			
White/European			
other/unidentified			

1B. General questions about the state assessment program

1.3 If any changes occurred in your state's assessment program during the 1995-96 school year, please describe below:

1.4 If any changes are planned for your state's assessment program, please describe below:

1.5 Please describe any steps your state takes to prevent test misuse or negative consequences from test use (steps could include state law or regulation, audits, investigations after allegations, training, etc.):

BEST COPY AVAILABLE

2. Professional Development

Person completing this section: Name____________________________________

Phone_______________ Fax _______________

2.1 Is professional knowledge of assessment required of or offered to teachers, administrators or other school personnel in your state on any of the topics listed below? Please respond using the following code:

P - required preservice;
I - required inservice;
O - offered by state;
S - available through state for teachers in a specific program or project;
N - offered in state, but not by state agency.

	Teachers	**School Administrators**	**Other School Personnel**
Traditional standardized tests			
Observational techniques			
Classroom assessment approaches			
Performance assessments			
Portfolios/ learning records			
State assessment programs			
Integrating classroom assessment and instruction			
Use of test results			
Psychometrics			

BEST COPY AVAILABLE

2.2 Does the state evaluate teacher competence in assessment? ___ Yes ___ No

If yes, describe how:

____

____

____

2.3 Does the state survey teachers and administrators to find out whether their professional development needs in assessment are being met? ___ Yes ___ No

If yes, please describe when and how, and attach reports:

____

____

____

2.4 How does the state evaluate the effectiveness of any materials or programs it presents to teachers or administrators for professional development in assessment?

____

____

____

3. State Assessment System Review

Please attach any reports you have available that respond to the questions in this sub-section.

Person completing this section: Name______________________________

Phone______________ Fax ______________

3.1 Has your state evaluated or surveyed assessment practices at the district, school or classroom levels?

Districts	___ Yes	___ No
Schools	___ Yes	___ No
Classrooms	___ Yes	___ No

Please attach a copy of any reports.

If you know of any similar evaluations or surveys about your state carried out by other persons/organizations, please provide a citation:

____

____

____

3.2 Does your state review the state's assessment system? ___ Yes ___ No

If no, skip to section 4; if only a component has been reviewed, skip to section 4.

3.3 Who is responsible for supervising the review of state's assessment system?
Name____________________ Phone _______________ Fax _______________

3.4 When was the assessment system last reviewed? ________________________

3.5 How frequently is it reviewed? ____

3.6 Was the last review conducted by (answer yes to both if appropriate):

the education department?
___ Yes ___ No If yes, how frequently______________________________

independent reviewers?
___ Yes ___ No If yes, how frequently______________________________

BEST COPY AVAILABLE

3.7 Does the review include studying the impact of assessment:

on curriculum? ___ Yes ___ No
on instruction? ___ Yes ___ No
on high school graduation rates? ___ Yes ___ No

3.8 Describe the involvement and role played, if any, of stakeholders and outside experts in evaluating the assessment system:

	Involved (Yes/No)	**Number involved**	**Describe role**
teachers			
administrators			
state ed dept: assessment staff			
content specialists			
special education			
LEP			
other___________			
other___________			
outside experts			
students			
parents			
general community			
ed. organizations			
advocacy/ community groups			
business groups			

4. State Assessment Program Components

For each program component, please answer the following questions in sub-sections A-F.

We have included one copy of section 4 of this survey for each assessment component you identified in the SSAPD, plus one extra copy of this section in case you have a new component (if you have more than one new component, please make a copy for each). A copy of the SSAPD glossary is appended to each copy of section 4.

We appreciate copies of materials about program components, including reports, samples from or whole assessments (we will keep them secure if requested), scoring guides, etc. If a question in Part 4 of this survey can best be answered by attaching and making reference to a document, please do so.

Name of program component:____

Contact person for component: Name______________________________

Phone____________ Fax ____________

4A. General Component Information

4.1 Does your state require this assessment component to be administered to:

private school students	___ Yes	___ No
homeschooled* students	___ Yes	___ No
public school students	___ Yes	___ No

*students educated by parents, not students at home or in hospital for illness

4.2 Does the state provide any preparatory material for students to practice on this component? ___ Yes ___ No

If yes, please describe the materials and the purposes they are to serve (e.g., familiarity with format, familiarity with typical content):

____

____

____

BEST COPY AVAILABLE

4.3 Does the state provide any professional development for teachers about this component? ___ Yes ___ No

If yes, describe briefly:

____

____

____

4.4 Does this component use item or student sampling procedures for any subject areas that are assessed with this component? Please check relevant boxes:

	multiple complete forms*	**matrix sample****	**grades at which students are sampled*****	**% students sampled**
math				
reading/ language arts				
writing				
science				
social sciences				
foreign language				
art				
other__________				
other__________				

* each form contains items that test all learning objectives to be tested, but the items are not identical;

** each form contains items that test only a sub-sample of the learning objectives to be tested (as in NAEP item blocks);

***if students within a grade level are sampled, enter the grade level(s) at which the sampling occurs.

4.5 If more than one assessment method is used in this component (e.g., multiple-choice and constructed-response short-answer), list the methods used, the proportion of assessment time spent on each method, and the proportion of the total component score allocated to each method. If this varies by grade or subject tested, give an average or a range:

____

____

____

____

____

4.6 Are descriptions of assessment methods, samples of assessments, scoring guides or rubrics, or examples of work of varying kind and quality distributed to:

	students	**teachers**	**administrators**	**parents**	**community**
methods					
samples					
guides					
examples					

4.7 Does the state assessment provide opportunity for students to comment or reflect on their learning (e.g., space in a portfolio or in a survey)? ___ Yes ___ No

If yes, please describe:

____

____

4.8 If a portfolio is used, who decides what work is included?
____ students
____ teachers
____ students and teachers together
____ other_____________________________
____ portfolio not used

If the decision varies by subject or grade level, note here:

____

____

4B. Assessment Uses and Consequences

4.9 How is this assessment component used for any of the decisions about ***students*** that are listed below? Please indicate:

- * whether it is used (Yes/No);
- * whether a cut score (enter yes or no) must be exceeded for a favorable decision (e.g., above 70 to graduate);
- * whether an alternative assessment (enter yes or no) can be accepted for use in the decision; if you answer yes, describe the assessment and the conditions under which it is acceptable;
- * if the component is not used for the decision, whether it will be used in the future (enter anticipated date);
- * if a cut score is not used, describe how the score is used together with what other information (including other assessment components) to make a decision.

Note: if necessary, use the back of this page to complete description.

	used	cut score	alternative	future	describe
grade promotion					
graduation from H.S.					
grade-level retention					
placement in a track/level					
gifted/talented placement					
special ed. placement					
bilingual placement					
other					

BEST COPY AVAILABLE

4.10 If this component is used to determine high school graduation or grade promotion, what is the maximum number of times a student can take this assessment component?

For graduation______________________
For promotion______________________

4.11 Are there students exempted from this component?

	Yes	No	% not tested on most recent administration
IEP/Disability			
LEP			
Parental option			
Other__________			

4.12 Can a student or parent/guardian appeal a score on the component?
____ Yes ____ No

4.13 Can a student challenge items on the component as being flawed? ___ Yes ___ No

4.14 Can students or parents review assessment items after completion of assessment?
___ Yes ___ No

If yes, describe process:

____

____

____

4.15 Does the state inform students about how assessment component results will be used? ___ Yes ___ No

If yes, describe how:

____

____

____

4.16 How is this assessment used for making decisions about ***schools***? Please indicate:

* whether this assessment component has any of the consequences listed below (Yes/No)?

* if it does not now have a given consequence, will it have such a consequence in the future (enter anticipated date of use)?

* if you answer "Yes" to any specific consequences, is the test the sole criterion -- is the test alone used to make a decision (enter Yes or No)?

* if it is not the sole criterion, describe how the decision is made, including the role of test scores (e.g., low test score *and* low attendance and graduation rates are combined to trigger a review which can lead to accreditation loss, takeover, or dissolution).

Note: if necessary, use the back of this page to complete description.

If you have additional information on school consequences, add it in the space for comment.

	Yes/No	Future	Sole	Describe
Funding gain				
Exemption from regulations				
Warnings				
Probation, watch lists				
Funding loss				
Accreditation loss				
Takeover				
Dissolution				
Other				

Comment: ____

____

____

4.17 Do you have studies of the consequences (e.g., impact on curriculum, instruction, grade retention, graduation, tracking, funding decisions, etc.) of using this component? ___ Yes ___ No

If yes, please describe (including who performed the studies):

____

____

____

____

____

4.18 Is there evidence that there have been positive or negative unintended consequences to using this component? ___ Yes ___ No

If yes, what have the consequences been?

____

____

____

____

4.19 Is the assessment component intended to guide curriculum and instruction?
___ Yes ___ No

4.20 If yes to 4.19, is there evidence that the results have been used as intended?
___ Yes ___ No

Describe:

____

____

____

4C. Who Participates in Assessment Development & Scoring

4.21 Who participated in designing the assessment component, writing items or tasks, writing scoring rubrics, selecting examples of work at various levels, scoring the component, or participating on a bias review committee? If known, fill in the number of people who participated; if this is not known, enter Yes or No.

	design	**items**	**rubrics**	**examples**	**scoring**	**bias**
teachers						
administrators						
state ed dept: assessment staff						
content specialists						
special education						
LEP						
other____________						
other____________						
outside experts						
students						
parents						
general community						
education organizations						
advocacy/ community groups						
business groups						

BEST COPY AVAILABLE

4.22 Using the same task categories as in question 4.21, indicate the percentage of participants in various demographic categories who are involved in (if for any of these categories, the percentage is not known, enter Yes or No):

	design/items/ rubrics/examples	scoring	bias review
African American			
American Indian			
Asian American/ Pacific Islander			
Latino/Hispanic			
White/European American			
Other/unknown racial/ethnic			
female			
male			
urban			
rural			
suburban			
low income			

4D. Equity Concerns

4.23 Has the assessment development process for this component attempted to take into account the variety of cultural backgrounds of the student population?
___ Yes ___ No

If yes, please describe what has been done:

____

____

____

4.24 Has the assessment development process for this component attempted to take into account the different learning styles exhibited by students? ___ Yes ___ No

If yes, please describe what has been done:

____

____

____

4.25 Have any assessments in this component for grade 3 or lower been reviewed for developmental appropriateness? ___ Yes ___ No

If yes, please describe, *and if possible, attach any reports.*

____

____

____

4.26 Are items/tasks pre-tested and analyzed for bias? ___ Yes ___ No

4.27 Are item/tasks analyzed after administration for bias? ___ Yes ___ No

4.28 Is there a bias review committee for this component? ___ Yes ___ No

If no, please skip to section 4E.

4.29 Describe generally the authority the bias review committee has, *or attach and refer to report or document*:

____

____

____

4.30 For what problems does the bias committee review the assessment (e.g., language, stereotypes, confusing context, etc.); please list, *or attach and refer to written guidance for review committee members*:

____

____

____

4.31 Does the committee have the authority to delete or modify items for the component?
___ Yes ___ No

4.32 How are members of the bias review committe selected?

____
____
____

4.33 Are there racial/ethnic composition requirements for the bias committee?
___ Yes ___ No

If yes, describe:

____
____
____

4E. Reporting

So that we can analyze the reporting done for this component, please attach and refer to copies of reports (in English) *based on or including this assessment component* that are given to students, parents, teachers and schools, state agencies, and the general public.

If answers to any of the questions in this section can be found in the attached reports, you can refer to the report and page number(s).

4.34 How soon after administration are assessment component results reported?
to students____________________
to parents ____________________
to schools ____________________
to the public ____________________

4.35 In what languages, other than English, are reports available?

____

4.36 Has the state conducted surveys or investigations:

to find out what assessment information parents or the public want reported?
___ Yes ___ No

to find out whether parents or the public understand the reports? ___ Yes ___ No

BEST COPY AVAILABLE

4.37 Does the state provide guidance on the use of the results? ___Yes ___No

If yes, to whom:
___ parents
___ students
___ teachers
___ school administrators
___ psychologists and counselors
___ district administrators
___ general public

4F. Review of Assessment Components

(This section is for component review; there is a separate section, 3, for the whole system.)

Please attach any reports you have available that respond to the questions in this sub-section.

4.38 Subsequent to initial administration, has this component been subject to any form of review, for any purpose other than bias or alignment with state content, learner, or performance standards, curriculum frameworks or curriculum? ___ Yes ___ No

If yes, describe:

____

____

4.39 Is this component intended to be aligned with state standards, curriculum frameworks, or state curriculum? ___ Yes ___ No

If no, skip to 4.43

4.40 Describe how the alignment between standards, curriculum framework or curriculum and the assessment component has been determined:

____

____

____

4.41 Has component alignment been evaluated? ___ Yes ___ No

If yes, by whom was the alignment evaluation done (check all that apply)?
___ State Education Department
___ Private Test Contractor for component
___ Other independent evaluator -- specify ______________________

BEST COPY AVAILABLE

4.42 Are there, within any academic subject for which there are standards, frameworks or curriculum, parts of the standards or framework or curriculum which are not tested ("parts" means areas of content, or academic skills such as ability to integrate, synthesize, or use knowledge)? ___ Yes ___ No

If yes, please describe:

____

____

____

4.43 Has this component, or items or tasks within it, been evaluated for how well or the extent to which the component elicits and assesses level of cognitive demand or complexity, or critical thinking, in the domain(s) assessed? ____ Yes ___ No

If yes, describe:

____

____

____

4.44 Do you have technical studies on this component? ___ Yes ___ No

If yes, describe or attach and make reference to:

____

____

____

4.45 Is there any process to revise the assessment component based on any of these studies? ___ Yes ___ No

If yes, describe:

____

____

____

Have any revisions been completed? ___ Yes ___ No

5. Comment

Please use this space to comment on this survey. We are interested in a) any additional information you think is important, and b) feedback on this questionnaire, either in general or for any specific sections or items.

If we were to administer a similar questionnaire in the future, are there specific items which should be eliminated? added? revised?

Thank you for your assistance.

BEST COPY AVAILABLE

Glossary for Use with Association of State Assessment Programs Annual Survey

Cloze procedure: a kind of assessment item that uses any of a variety of fill-in-the-blank procedures, where the blank is embedded in a textual context.

Component: For the purposes of this survey, determine the number of components your state has in its total assessment program according to : 1) the form of the assessment used, and 2) the way results are used. For example:

1) Different formats: If your state has a criterion-referenced test, a norm-referenced test, and a writing sample, you would report three components; or
2) Different purposes: If your state uses an assessment primarily to determine high school graduation, another to assess school readiness, and a third to determine student's achievement compared to a state standard, you would report three components.

However, if your state uses one test for several different purposes, report only one component. On the other hand, if you use a number of different formats (for example, portfolio, NRT and CRT) for one purpose (for example, high school graduation), report three components. Finally, do not report separate components if the only difference between components is the subject area covered. For example, if your state uses a norm-referenced test in reading, mathematics, social studies, and science, report only one testing component.

Computer-adaptive testing: any assessment that requires the student to respond to the assessment items or task with the aid of a computer where the software selects next problem or task based on the student's prior responses.

Content standards: statements which specify what students should know or be able to. When set by states, these statements tend to be general and less concrete than performance standards.

Curriculum frameworks: one mechanism for linking learner standards and state goals. These frameworks provide sufficient guidance to curriculum developers and teachers throughout a state to ensure that curriculum and instruction drive towards the state goals while assuring that learner standards are met. Examples are Indiana's Proficiency Guides and California's Curriculum Frameworks.

Enhanced multiple-choice: any multiple-choice question that requires more than the selection of one correct response. Often, the task requires the students to explain their responses.

Extended-response, open-ended: any item or task that requires the student to produce an extended written response to an item or task that does not have one right answer (for example, an essay or laboratory report).

FTE: Full time equivalent.

Final Title I Assessment Plan: The final Title I plan for assessment or evaluation of student performance which states must submit by the 2000-2001 school year will need to meet all of the requirements in Improving America's Schools Act.

Group performance assessment: any assessment which requires the students to perform the assessment task in a group setting. For example, a performance assessment, as defined in individual performance assessment, becomes a group performance assessment when the task is performed in a group and the individual's rating is based on his performance as part of that group.

Individual performance assessment: any assessment that requires the student to perform (in a way that can be observed) an assessment task by him- or herself. For example, students may be asked to perform a laboratory experiment or carry out a community service project, and write up results. The performance of the laboratory experiment and the community service project makes this an individual performance assessment vs. an extended-response when the quality of the performance itself, and not just the quality of the writing is rated.

Interview: an assessment technique where the student responds to verbal questions from the assessor.

LEA: local education agency; refers to the school district.

Learner standards: statements which specify what students should know or be able to do. When set by states, these statements tend to be general and less concrete than performance standards. An example would be, "Students in our state shall write in a variety of forms, e.g., notes, letters, instructions, stories, and poems, for a range of purposes, e.g., to plan, inform, explain, entertain."

Measures of the enacted curriculum: the presence of educational approaches necessary to provide students with appropriate instruction on which they will be assessed; "opportunity to learn" standards hold the school accountable for providing these learning opportunities.

Non-traditional test items: any assessment activity other than a multiple-choice item from which the student selects one response. These items or performances are scored or rated using an agreed-upon set of criteria which may take the form of a scoring guide, a scoring rubric, or comparison to benchmark papers or performances.

Observation: an assessment technique that requires the student to perform a task while being observed and rated using an agreed-upon set of scoring criteria.

Opportunity to learn standards: see "Measures of the enacted curriculum."

Performance Standards: how well a student has to perform in order to perform at a satisfactory or other specified level.

Portfolio: an accumulation of a student's work over time which demonstrates the student's best performance, typical performance, or growth in performance.

Project, exhibition, or demonstration: a complex task over time which requires the demonstration of the mastery of a variety of desired standards, each with its own performance criteria, which can be assessed within the one project, exhibition, or demonstration.

SEA: State education agency.

Short-answer, open-ended: any item or task that requires the production of a short written response on the part of the respondent. most often, there is a single right answer (for example, a fill-in-the-blank or short written response to a question). "Constructed" response items, where the student grids the answer directly (not picking from a list) are included in this definition.

State goals: statements which specify desired or valued expectations for students, schools, or school systems. They do not say what students should know or what schools should do. They do detail the end-points of the educational enterprise, the reasons schools exist. An example would be, "All people of this state will be literate, lifelong learners who are knowledgeable about the rights and responsibilities of citizenship and able to contribute to the social and economic well-being of our diverse, global society."

Student expectations: statements which specify what students should know or be able to. When set by states, these statements tend to be general and less concrete.

Transitional Title I Assessment Plan: The Title I assessment and evaluation plan states will use between 1995-96 and the 2000-2001 school years to assess the impact of Title I programs on students.

Appendix E

FAIRTEST ASSESSMENT SURVEY -- SHORT FORM

State ____

Respondent name, title, phone

____

____

1. System information

1.1. List state assessment components, including for each the methodologies used (e.g., multiple-choice, constructed response, etc.), whether it is norm- or criterion-referenced, the subjects and grades assessed.

1.2. If any changes are planned for your state's assessment program, please describe.

1.3 Has the state conducted surveys or investigations:

to find out what assessment information parents or the public want reported?

to find out whether parents or the public understand the reports?

___ Yes ___ No

2. Professional Development

2.1. Does your state have any requirements for pre-service training in assessment for teachers? If yes, what kinds of knowledge is required (e.g., state tests, traditional standardized tests, classroom assessment, performance assessment).

2.2 Does the state evaluate teacher competence in assessment? ___ Yes ___ No
If yes, describe how:

2.3 Does the state survey teachers and administrators to find out whether their professional development needs in assessment are being met? ___ Yes ___ No

2.4 How does the state evaluate the effectiveness of any materials or programs it presents to teachers or administrators for professional development in assessment?

3. State Assessment System Review

3.1 Has your state evaluated or surveyed assessment practices at the district, school or classroom levels?

Districts ___ Yes ___ No
Schools ___ Yes ___ No
Classrooms ___ Yes ___ No

Please attach a copy of any reports.

If you know of any similar evaluations or surveys about your state carried out by other persons/organizations, please provide a citation:___________

3.2 Does your state review the state's assessment system? ___ Yes ___ No

3.3 When was the assessment system last reviewed? _________

3.4 How frequently is it reviewed? ________

3.5 Was the last review conducted by (answer yes to both if appropriate):
the education department? ___ Yes ___ No
independent reviewers? ___ Yes ___ No

3.7 Does the review include studying the impact of assessment:
on curriculum? ___ Yes ___ No
on instruction? ___ Yes ___ No
on high school graduation rates? ___Yes ___ No

4. State Assessment Program Components

For each item, note which component a yes response applies to.

4.1 Does the state provide any preparatory material for students to practice on?____

4.2 Does the state provide any professional development for teachers? ___
If Yes, describe briefly for relevant components.

4.3 Are there students exempted?

IEP:

LEP:

Parental option:

4.4 Can a student or parent/guardian appeal a score?

4.5 Can a student challenge items as being flawed?

4.6 Can students or parents review assessment items after completion of assessment?

If yes, describe process:

4.7 Does your state use a test as a high school graduation requirement? (Name of test).

Does your state use a test as a requirement for grade promotion? (name)

If yes to either, are students with an IEP required to pass the test?

students with LEP?

For any student who does not pass the test, is any alternative available that would allow the student to obtain the same diploma as if s/he had passed the test, or be promoted? Please describe.

4.9 Does this assessment have consequences for schools?

If yes, is it the sole basis or one factor for the consequences?

Description/comment

4.10 Do you have studies of the consequences (e.g., impact on curriculum, instruction, grade retention, graduation, tracking, funding decisions, etc) of using any component?

If yes, please describe (including who performed the studies)

4.11 Is there evidence that there have been positive or negative unintended consequences to using any component? ___ Yes ___ No

If yes, what have the consequences been?

4.12 Is the assessment component intended to guide curriculum and instruction?

If yes, is there evidence that the results have been used as intended?

Describe:

4.14 Has the assessment development process attempted to take into account the variety of cultural backgrounds of the student population?

If yes, please describe what has been done:

4.15 Has the assessment development process for this component attempted to take into account the different learning styles exhibited by students?

If yes, please describe what has been done:

4.16 Have any assessments in this component for grade 3 or lower been reviewed for developmental appropriateness?

4.17 Are items/tasks pre-tested and analyzed for bias?

4.18 Are item/tasks analyzed after administration for bias?

4.19 Is there a bias review committee?

Does the committee have the authority to delete or modify items?

What is the racial/ethnic composition of the committee.

How closely does this reflect the state population?

the student population?

4.20 Are public reports about assessment available in any language(s) other than English? If yes, list.

4.21 Are any components intended to be aligned with content, learner, or performance standards, curriculum frameworks, or state curriculum?

If no,stop here

4.23 Describe how the alignment between standards, curriculum framework or curriculum and the assessment component has been determined:

4.24 Has component alignment been evaluated? ___ Yes ___ No

If yes, by whom was the alignment evaluation done (check all that apply)?

___ State Education Department
___ Private Test Contractor for component
___ Other independent evaluator -- specify _______

4.25 Are there, within any academic subject for which there are standards, frameworks or curriculum, parts of the standards or framework or curriculum which are not tested? ("parts" means areas of content, or academic skills such as ability to integrate, synthesize, or use knowledge) ___ Yes ___ No

If yes, please describe:

4.26 Has this component, or items or tasks within it, been evaluated for how well or the extent to which the component elicits and assesses cognitive complexity or critical thinking in the domain(s) assessed? ____ Yes ___ No

If yes, describe:

4.27 Do you have technical studies on this component? ___ Yes ___ No

If yes, describe or attach and make reference to:

4.28 Is there any process to revise the assessment component based on any of these studies? ___ Yes ___ No

If yes, describe:

Have any revisions been completed? ___ Yes ___ No

Appendix F

Excerpts (pp 4-19) from

Principles and Indicators for Student Assessment Systems

by the

National Forum on Assessment

Published by FairTest, 1995

Educational Foundations for High Quality Assessment

Developing the Principles and Indicators *required the Forum to define underlying beliefs and identify the essential conditions that enable high quality schooling. These form a foundation for high quality assessment systems.*

The Forum agrees on the following beliefs:

- All students deserve the opportunity to learn high-level content in and across subject areas and to learn in a resource-rich, supportive environment.
- Thinking is the most basic and important skill.
- High achievement takes many forms.
- Equity demands equivalence in the standards of learning for all students and in the instructional quality offered to each student, together with the opportunity to demonstrate learning in a variety of ways.
- Family and community support is essential to student success.

The Forum views the following four conditions as necessary for schools to ensure successful learning and support the assessment practices promoted by the Principles:

1. ***Schools organize to support the multiple learning needs and approaches of all their members.***

Schools foster a supportive environment for inquiry, intellectual challenge, and cooperation. The school climate and professional development for teachers and administrators promote respect for and inclusion of females and males from all ethnic, disability, language, socio-economic, and cultural groups. The school works toward the elimination of racism, sexism, and bias. It provides a safe environment for all students. It democratically involves all its members in shaping the school's learning and governing life, while recognizing that students require guidance in their growth to adulthood and independent learning and that students, educators, families, and support staff have different roles in assuring student success.

The school recognizes that learning is not housed in just one building. It develops collaborative external relationships so that students interact with and learn from members of the wider community, who, in turn, are welcomed by the school.

The school continually evaluates itself in order to improve. Assessment focuses on providing information used to strengthen student learning and on documenting progress. The school helps prepare educators to evaluate all students fairly. Assessments provide useful information on the particular knowledge and abilities students have or have not yet developed, in ways that will guide further learning and the improvement of curriculum and instruction.

2. ***Schools work to understand how learning takes place and what facilitates learning.***

Learning is an intellectually active and social process shaped by the learner's experiences, perceptions, and culture. Schools provide the environment, curriculum, and instruction to facilitate active learning by both students and educators. Educators use new knowledge on learning to improve teaching and assessment.

Because learning requires feedback and reflection, assessment is an essential component of the process. To be helpful, an assessment system uses methods that are compatible with how different students learn, provides information on how each student learns, and offers a variety of methods and opportunities for demonstrating achievement.

3. ***Schools establish clear statements of desired learning for all students and help all students achieve them.***

Such statements are also called learning goals or content standards. They describe broad, important intellectual competencies—knowledge, skills, understandings, and habits of mind—that students should acquire and be able to demonstrate. These include important learning in and across subject areas, with a focus on thoughtful application and meaningful use of knowledge.

In order to establish general public agreement, statements of desired learning are determined through open discussion among subject-matter experts, educators, families of students, policymakers, students, and other members of the wider community, including advocacy, business, higher education, and civic organizations.

Assessment systems rely on practices and methods that are integrated conceptually with curriculum and instruction which, in turn, are based on the statements of desired learning. Schools use assessments to help students learn as well as to document and evaluate their learning.

4. ***All schools have equitable and adequate learning resources and classroom conditions, including capable teachers, a rich curriculum, safe and hospitable buildings, sufficient equipment and materials, and essential support services.***

Taken together, these conditions provide an opportunity to learn. Class sizes are small enough that teachers are able to get to know and work closely with all their students and use active approaches to learning and assessment. Tracking and full-time, long-term placements out of the mainstream classroom generally do not occur, but if determined necessary, are periodically assessed for effectiveness. Teachers have sufficient time to plan learning and assessment activities, discuss student learning, and work with fellow teachers. They have access to adequate professional development resources.

Reports to the public on student learning include valid and coherent information on available learning resources and conditions. This is necessary in order to help evaluate any impact resources have on learning and to facilitate obtaining needed resources if they are absent. It also helps create a climate in which students are not held responsible for the absence of equitable or adequate resources.

•••

This picture is ideal. It provides a vision of excellent education for all students to which good assessment makes a vital contribution. Assessment reform and broader school reform can and should move forward together.

PRINCIPLE 1 The Primary Purpose of Assessment Is to Improve Student Learning

Assessment systems, including classroom and large-scale assessment, are organized around the primary purpose of improving student learning.

Assessment systems provide useful and accurate information about student learning. They employ practices and methods that are consistent with learning goals, curriculum, instruction, and current knowledge of how students learn. Educators assess and document student learning through an appropriate balance of methods that can include structured and informal observations and interviews, projects and tasks, experiments, tests, performances and exhibitions, audio and video tapes, portfolios, and journals. The consequences of using an assessment or a particular method are evaluated regularly to ensure that its effects are, in fact, educationally beneficial.

Classroom assessment is the primary means through which assessment affects learning. It is integrated with curriculum and instruction so that teaching, learning and assessing flow in a continuous process. By documenting and evaluating student work over time, teachers obtain information for understanding student progress in ways that can guide future instruction. Assessment also provides opportunities for self-reflection and evaluation by the student.

Teachers are the primary users and developers of classroom assessments. They understand and apply, as appropriate for classroom work, current technical concepts of effective assessment practices, particularly validity and reliability. Individually and in groups, they analyze the impact of different assessments on student learning and use the results of their analyses to improve their assessment practices.

For classroom and large-scale assessments, scoring guides ("rubrics") for evaluating student work are stated in positive terms (what a student can do) and are appropriate to the work being done. They present a coherent picture of how students can develop and improve their performance.

No assessment method or practice is used that narrows or distorts the curriculum or instructional practices. Multiple-choice and short-answer methods, if used, constitute a limited part, in time or impact, of the total assessment system. History shows that their use, if too prominent, can skew instruction away from methods of teaching that support important learning.

In documenting student achievement, systems focus on providing information grounded in clearly defined learning goals for students and information about a student's progress. Therefore, assessments intended to rank order students or compare students with each other are not a significant part, in time or impact, of the total assessment system.

Principle 1: Indicators

1. Assessments are based on curriculum and desired learning outcomes that are clearly understood by students, educators, and parents.
2. Assessment practices are compatible with current knowledge about how learning takes place and allow for variety in how students learn.
3. Assessment systems enable a process of continuous feedback for the student.
4. Most assessments allow students to demonstrate understanding by thoughtfully applying knowledge and constructing responses.
5. Assessment systems allow students multiple ways to demonstrate their learning.
6. Assessment systems include opportunities for individual and group work.
7. Classroom assessments are integrated with curriculum and instruction.
8. Teachers employ a variety of assessment methods and obtain multiple forms of evidence about student learning for planning and implementing instruction and for evaluating, working with, and making decisions about students.
9. Teachers can explain how their assessment practices and instruments help improve teaching and how they provide useful information for working with students.
10. Student self-reflection and evaluation are part of the assessment system.
11. Schools establish procedures for enabling classroom-based student assessment information to follow each student from year to year.
12. Assessment methods, samples of assessments, scoring guides or rubrics, and examples of work of varying kind and quality are discussed and understood by students.
13. Scoring guides (rubrics) state in positive terms what students can do and enable users to analyze student strengths and needs in order to plan further instruction.
14. Educators make clear to students the uses and consequences of each assessment.
15. Teachers use current principles and technical concepts of assessment, particularly validity and reliability, in developing and analyzing their classroom assessments.
16. Multiple-choice and short-answer methods are a limited part, in time or impact, of the total assessment system.
17. Assessments intended to rank order students or compare students with each other are not a significant part, in time or impact, of the total assessment system.

PRINCIPLE 2 Assessment for Other Purposes Supports Student Learning

Assessment systems report on and certify student learning and provide information for school improvement and accountability by using practices that support important learning.

In order to support learning, assessment for these purposes conforms to the spirit and general requirements of Principle 1. When teachers, schools, districts, and states all use assessment practices and methods which are consistent with learning goals and current knowledge of how students learn, they establish the basis for a coherent system which meets a variety of purposes.

To report student learning to families, students and other educators, to certify student achievement, or to make important educational decisions, teachers analyze assessment information from ongoing school work and assessments. Important decisions about individuals, such as program placement, grade promotion, or graduation, are not made on the basis of any single assessment.

To provide information useful for school improvement, teachers and other school staff primarily rely on assessment information that is based on regular, continuing work by the school's students. External or large-scale assessments provide additional and corroborative information.

To provide information for accountability, the school, the district, and the state gather a variety of assessment information that they can use to inform the public, provide assistance to schools and districts, and make decisions about programs. This information can come from a combination of classroom-based assessment information (such as portfolio reviews) and external or large-scale assessments (such as examinations). To evaluate programs efficiently, districts and states rely on various forms of sampling, to the extent feasible. Technical standards for assessment are revised or developed to ensure they are adequate for the assessment purposes and methods, and they are used to help ensure high quality practices. Research is conducted to ensure that assessments are supporting and not harming important student learning. Because the context of learning affects student achievement and all students are held to the same high standards, accountability reports include contextual information about resources, school practices and quality, and other outcomes.

Principle 2: Indicators

1. Teachers, schools, districts, or states make reports on and decisions about individuals on the basis of cumulative evidence of learning, using a variety of assessment information, not on the basis of any single assessment.
2. Assessment systems provide students with multiple opportunities to demonstrate their learning.
3. Schools use assessment information to improve curriculum, instruction, and teacher effectiveness.
4. The evaluation of an accumulation of work and assessments done by students over time is a major component of accountability.
5. Information for accountability is obtained through sampling, to the extent feasible.
6. When classroom-based information is used in accountability, independent evaluations of the information, such as re-scoring a sample of the portfolios or exams, are conducted.
7. Teachers view assessments for accountability purposes as consistent with and not harmful to curriculum, instruction, and high quality classroom assessment.
8. Information from large-scale assessments is returned to the school and teachers in a form that they can use.
9. If programs, schools, districts, or states are compared, appropriate contextual information is provided.
10. Technical standards for assessment systems are developed and used to ensure that assessments provide accurate and comprehensive information, measure progress toward learning goals in ways that are consistent with how students learn, and are used appropriately.
11. Technical studies of large-scale assessments or those used across a number of classrooms or schools show that the assessments focus on important knowledge as defined in learning goals, are consistent with knowledge of how students learn, and are not biased against particular population groups.
12. Validity studies of large-scale assessments or those used across a number of classrooms or schools show that the assessments have beneficial, not harmful, effects on student learning and that actions taken based on assessment information are adequately supported by and are appropriate uses of that information.

PRINCIPLE 3 Assessment Systems Are Fair to All Students

Assessment systems, including policies, practices, instruments, and uses, are fair to all students.

Assessment systems ensure that all students receive fair treatment in order not to limit students' present education and future opportunities. Assessment is fair when every student has received equitable and adequate schooling, including culturally sensitive curriculum, instruction, and assessment that encourage and support each student's learning, and when assessment systems meet these *Principles*. In particular:

Assessment results accurately reflect a student's actual knowledge, understanding and achievement. Assessments are designed to minimize the impact of biases on the student's performance, including:

- biases of persons developing or conducting the assessment, evaluating the performance, or interpreting or using the results;
- biases caused by basing assessments on the perspectives or experiences of one particular group; and
- biased format or content, including offensive language or stereotypes.

Educators and assessment and content experts construct assessment systems that support learning by all students in a diverse population with varying learning styles. Assessment developers and users recognize and build upon the benefits of diversity. Assessment systems allow for multiple methods, as stated in Principle 1, to assess student progress toward meeting learning goals and for multiple but equivalent ways for students to express knowledge and understanding. Assessments are administered under conditions that support high quality performance.

Assessments are created or adapted and accommodations are made to meet the specific needs of particular populations, while preserving the integrity and validity of the assessments. These populations include English language learners (also identified as limited English proficient students) and students with disabilities. Adaptations include, but are not limited to, physical accommodations, assessments in a student's primary language or language of instruction (written, oral or signed), and extra time. Advocates for specific groups help detail how to meet these assessment standards.

Students should not suffer adverse consequences simply because their backgrounds or school experiences may have made them less familiar with particular methods of assessment. Therefore, teachers and schools provide all students with instruction and practice in the assessment methods used to evaluate their progress, but do not engage in inappropriate coaching.

Assessment developers consider possible adverse consequences of using the assessment, particularly for those groups which currently suffer discrimination or the effects of previous discrimination. Assessments are modified as necessary to reduce harmful impacts while preserving accuracy. Assessments are used to provide students with optimal learning opportunities, rather than place them in tracks or programs which narrow curriculum options or foreclose educational opportunities.

Principle 3: Indicators

1. Every student has the opportunity to perform on a variety of high quality assessments during the school year.
2. Schools prepare all students to perform well on assessments which meet these principles.
3. Assessment practices recognize and incorporate the variety of cultural backgrounds of students who are assessed.
4. Assessment practices incorporate the variety of different student learning styles.
5. Assessments, particularly for young children, are developmentally appropriate.
6. Assessments are created or adapted to meet the needs of students who are learning English.
7. Assessments are created or adapted and accommodations made to meet the needs of students who have a disability.
8. All students are knowledgeable and experienced in the assessment methods used to evaluate their work.
9. The group which designs or validates an assessment reflects, has experience with, and understands the particular needs and backgrounds of the student population, including race, culture, gender, socio-economic, language, age, and disability status.
10. Committees of persons knowledgeable about the diverse student population review large-scale assessments for bias and are able to modify, remove, or replace items, tasks, rubrics, or other elements of the assessment, if they find them biased or offensive.
11. Teacher education and continuing professional development prepare teachers to assess all students fairly.
12. Technical standards are developed and used to ensure that assessments do not have harmful consequences for student learning or teaching.
13. States and districts report their assessment data by racial, ethnic, gender, linguistic, disability, and socio-economic status groups for analysis of school, district, and state results, provided that doing so does not infringe upon student privacy rights.
14. Schools do not use assessments to track or place students in ways that narrow curriculum options or foreclose educational opportunities.

PRINCIPLE 4 Professional Collaboration and Development Support Assessment

Knowledgeable and fair educators are essential to high quality assessment systems and practices.

Assessment systems depend on educators who understand the full range of assessment purposes, use appropriately a variety of suitable methods, work collaboratively, and engage in ongoing professional development to improve their capability as assessors.

Teachers are the primary assessors. They:

- document, evaluate, and report student learning;
- construct, select, and use appropriate, high-quality methods and instruments to meet various assessment purposes; and
- participate in developing and scoring any district or state assessments and know how to use relevant information from them.

Schools of education assess their own students using methods they expect prospective teachers to learn to use. They prepare administrators to support, assist, and supervise teachers in high quality assessment practices. They prepare teachers to:

- integrate assessment with instruction and curriculum;
- use a variety of high quality methods for assessing the performance and development of a diverse student population; and
- communicate with families and students about the methods used and the information obtained from the assessments.

Educators, including teachers, administrators, teacher aides, school psychologists, and counselors, determine the types of individual and collective professional development that contribute to the quality of assessment practices. They actively participate in such professional development. They work together to improve their craft, meet regularly to discuss assessment and evaluate student work, and establish networks to discuss assessment issues and practices, particularly in the fields they teach. They engage in scoring and discussing portfolios, work samples, or performance examinations at the district or state level. They consult with families, the community, and various experts to shape professional development in assessment to meet the needs of all their students. Schools, states, and districts provide resources that educators can call on or use as appropriate to strengthen their assessment capabilities.

Principle 4: Indicators

1. Teacher educators ensure that beginning teachers possess basic knowledge, skills, and experience for assessing their students with a variety of appropriate methods and communicating with parents and students.
2. Teacher educators practice appropriate assessment techniques.
3. Teachers perform well in their role as primary assessors of student learning.
4. Teachers regularly participate in setting performance standards, selecting examples of work of different quality, and scoring or re-scoring portfolios or performance assessments at the school, district, or state level.
5. Teachers and administrators know how to use the results of large-scale assessment information for program and school improvement.
6. Schools and districts provide regular, substantial meeting time for collaborative professional development that includes discussions of assessment, actual student work, and the relationship of assessment to instruction.
7. Educators work together to determine the professional development needed for improving their capabilities as assessors.
8. Educators actively participate in professional development for improving their capabilities as assessors.
9. Teachers and other school personnel consult with parents and other community members about professional development related to assessing all students in the school.
10. Schools, districts, and states provide adequate opportunities for administrators to engage in professional development that supports sound teacher and school assessment practices.
11. Schools and districts enable teacher aides, counselors, psychologists and other school personnel to participate with teachers, as appropriate, in professional development about assessment.
12. Districts and states provide resources needed for professional development.

PRINCIPLE 5 The Broad Community Participates in Assessment Development

Assessment systems draw on the community's knowledge and ensure support by including parents, community members, and students, together with educators and professionals with particular expertise, in the development of the system.

Parents, family members, and students contribute important information and knowledge to both classroom and large-scale assessments. This includes knowledge about how students learn, the communities and cultures in which they live, and how children can be prepared for assessment experiences. School systems educate family and community members to participate effectively in the assessment system and provide information about how parents can support their children in the assessment process. School systems also educate parents and the community about the meaning of assessment results. Schools, districts and other assessment developers create a supportive atmosphere, ensure accessible meeting times and places, and use language that encourages broad-based community participation in planning, designing, and evaluating the assessments.

In constructing, selecting, and using assessments for their classrooms, teachers incorporate and build on parent, family, community, and expert knowledge. Developers of large-scale assessments include teachers and other school-based educators in the development process.

Assessment, curriculum, and content experts continue to have a central role in developing large-scale assessments. They also have a responsibility to help teachers and schools develop and improve classroom assessment practices. Experts are particularly attuned to teachers' needs to improve assessments within the everyday constraints and challenges of teaching. Teachers and administrators, in turn, consider the insights provided and issues raised by the experts.

Other evaluators of students, such as counselors and psychologists, work with teachers, relying primarily on analysis of classroom activity to plan how best to educate each child.

Principle 5: Indicators

1. Teachers, schools, districts, states, and other assessment developers include students, family, and community members in planning, developing, reviewing, and evaluating assessment systems, instruments, and practices.
2. Schools and districts educate parents and community members to participate effectively in developing and reviewing assessment systems and practices.
3. Teachers, schools and districts educate parents and community members about the meaning and interpretation of assessment results.
4. Those developing assessments ensure that meeting times and places are accessible to all people who desire to participate in assessment development.
5. Schools and teachers provide parents the opportunity to discuss classroom assessment practices.
6. Students participate in discussing standards and planning both classroom and large-scale assessments.
7. Teachers, school administrators, and other school personnel from a variety of subject areas, grade levels, and demographic backgrounds play a prominent role in designing, administering, and scoring any assessments mandated by the school, district, state, or federal government.
8. Assessment, curriculum, and content experts work together with school-based educators to develop assessments that support important learning, are compatible with how students learn, and promote effective instruction.

PRINCIPLE 6 Communication about Assessment Is Regular and Clear

Educators, schools, districts, and states clearly and regularly discuss assessment system practices and student and program progress with students, families, and the community.

Educators, schools, districts, and states communicate, clearly and in ordinary language, the purposes, methods, and results of assessment. They focus their reporting on what students know and are able to do, what they need to learn to do, and what will be done to facilitate improvement in learning. They report achievement data in terms of learning standards and avoid comparing students or programs in ways that do not support good instructional practices. Teachers and schools also clearly inform parents and students about important assessments, including what the assessment is, when it will occur, and how the results will be used.

Schools, districts, and states make use of many avenues of communication (with appropriate protection for student privacy), including parent-teacher conferences, mass media, school papers, displays of student work in public spaces, and open meetings to view and discuss student work and assessment results. They also provide translations (written, oral, or signed) of important information into languages used by the families and communities served. Information on all students in the system is included in public reports by schools, districts, and states.

Teachers, schools, districts, and states establish avenues for comment and feedback from family and community members about the assessment processes. Educators and technical experts work with families and communities to improve reporting and plan how best to receive and use feedback to improve assessment practices. Specialized or technical information intended primarily for professional use is also readily available to the public.

Schools, districts, and states present assessment results in conjunction with other information about schooling, including information about:

- education programs, including curriculum, instructional practices, student placement practices, and class size;
- social data, including poverty indices and demographic data on students, staff, and community;
- resources, including funding and expenditures, staff qualifications, and available materials and equipment;
- school environment, including building quality and freedom from violence; and
- outcomes, including graduation rates, post-secondary education attendance, and other measures of long-term achievement and satisfaction.

Principle 6: Indicators

1. Survey results show that parents and other community members from different racial, ethnic, cultural, income, disability, and linguistic groups agree that reports:
    - are clear;
    - are sufficiently frequent; and
    - include sufficient examples of goals, standards, sample or actual assessments, rubrics or scoring guides, and examples of student products (with safeguards for privacy).
2. Parents, students, and other community members participate in determining the content, form, and frequency that reporting will take.
3. Translations enable all parents with limited or no English proficiency to receive information about the achievement of their children; and they enable all community members to receive data about student achievement in general at the school, district, and state levels.
4. Reports on schools, districts, or states include information on all students.
5. Schools, districts, and states report achievement information to the public in terms of agreed-upon learning standards.
6. Schools and teachers report individual student achievement information to students and families in terms of learning standards, individual growth and progress, student interests, and how the student learns.
7. School and teacher reports about student achievement focus on what students know and are able to do, what they need to learn to do, and what will be done to facilitate improvement.
8. Teachers and schools present information in a variety of ways, including written reports and conferences, to students and their families.
9. Teachers clearly inform students and parents about important assessments, including what, when, and how they are used.
10. Schools, districts, and states use many avenues of communication to inform the public.
11. All reports explain the meanings, limitations, and strengths of reported data.
12. Public reports present assessment information in the context of education programs, social data, resources, school environment, and other outcomes.
13. Technical and specialized reports are readily available to interested members of the public.

PRINCIPLE 7 Assessment Systems Are Regularly Reviewed and Improved

Assessment systems are regularly reviewed and improved to ensure that the systems are educationally beneficial to all students.

Assessment systems must evolve and improve. Even well-designed systems must adapt to changing conditions and increased knowledge. A periodic, comprehensive review is the basis for making decisions to alter all or part of the assessment system. In this review process, educators use these *Principles and Indicators*, including the "Foundations" section. An assessment review usually is integrated with a review of the educational system as a whole.

The ultimate value of an assessment system is its ability to enhance learning for all students. Reviews involve an inquiry process focused on two questions: Does the system provide information useful for making decisions and taking action? Are the actions taken educationally beneficial?

Reviewers consider how well the information provided by assessments helps in making decisions and improving schooling. They pay careful attention to any unintended consequences of the assessment system, particularly on teaching and learning, and especially for groups who suffer discrimination or the effects of previous discrimination. Reviewers consider how well the system adheres to each of the assessment principles. They also consider how well the parts of the assessment system combine to form a coherent whole. If only part of a total system is reviewed (e.g., one school's assessments), the review is tailored to fit the purposes of that part.

To ensure that timely and effective reviews are conducted, a continuing group has responsibility for monitoring the review process. The primary reviewers of classroom assessments are school-based educators working collaboratively. Parents, students, and other educators and experts also provide feedback about classroom and school practices. Assessment reviews by schools are part of regular evaluations of school quality. Reviews of large-scale assessments and whole systems require broad participation from all stakeholder groups, including teachers and other educators; family and community members; advocacy, civil rights, higher education, business, labor and community groups; students; and assessment and curriculum specialists. Independent expert analysis of the system is included in the public review process.

Reviews include an analysis of the costs and benefits of the assessments to the education system as a whole. The most important criterion for cost-benefit analysis is that the assessment benefit and not harm important student learning.

Schools, districts, and states use review information to improve the system. Because new programs or fundamental changes take time to show results, school systems do not use assessment review information to make hasty decisions about programs; nor do they use difficulties in implementing new assessments that are consistent with these *Principles* as a reason to quickly discard them.

Principle 7: Indicators

1. The assessment system at all levels is reviewed regularly.
2. A continuing group has responsibility for monitoring the assessment review process.
3. Surveys show that stakeholders were able to participate in evaluating school, district, and state assessment systems.
4. Public review of the assessment system includes analysis by independent experts in curriculum, instruction, and assessment.
5. Cost-benefit analyses of the assessment system focus on its effects on instruction and learning.
6. The review includes evidence of the use of assessment information in the educational planning and improvement process.
7. Reviewers evaluate:
    - adequacy of classroom assessment practices tc support important learning for all students;
    - effects of assessments on curriculum, instruction, and learning;
    - adequacy of information for certification, program improvement, and accountability;
    - fairness for all students;
    - technical quality and rigor of assessments;
    - intended and unintended consequences of the assessment system, particularly those affecting learning and equity;
    - adequacy of professional development activities;
    - extent and quality of professional collaboration on assessment;
    - extent and quality of stakeholder involvement in developing and reviewing the assessment system;
    - adequacy of contextual information that is presented with assessment data and used to help understand student learning outcomes;
    - quality of communication with families and the public;
    - costs and benefits of the assessment system;
    - quality and usefulness of the review process itself; and
    - coherence of the assessment system.

Appendix G

Bibliography

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1985). *Standards for Educational and Psychological Testing.* Washington, DC: American Psychological Association.

American Federation of Teachers (AFT). (1996). *Making Standards Matter: 1996.* Washington, DC: AFT.

Council of Chief State School Officers (CCSSO). (1996). *States' Status on Standards: 1996 Update.* Washington, DC: CCSSO. (Note this report actually covers 1994-95.)

Council of Chief State School Officers and North Central Regional Educational Laboratory (CCSSO/NCREL). (1997). *Annual Survey of State Student Assessment Programs, 1995-96.* Washington, DC: CCSSO.

Council of Chief State School Officers and North Central Regional Educational Laboratory (CCSSO/NCREL). (1997). *Trends in State Student Assessment Programs, 1995-96.* Washington, DC: CCSSO.

Council of Chief State School Officers and North Central Regional Educational Laboratory (CCSSO/NCREL). (1996). *State Student Assessment Programs Database, 1994-95.* Oak Brook, IL: NCREL.

Council of Chief State School Officers and North Central Regional Educational Laboratory (CCSSO/NCREL). (1994). *State Student Assessment Programs Database, 1993-94.* Oak Brook, IL: NCREL.

Fairbanks, A. H., and Roney, C. E. (1995). *Assessing State Assessments: An Analysis of Movements Toward Performance Assessment.* Cambridge, MA: Unpublished manuscript submitted to FairTest as part of coursework at the John F. Kennedy School of Government, Harvard University.

FairTest. (1996). *Performance Assessment: Annotated Bibliography and Resources* (third ed.). Cambridge, MA: FairTest.

FairTest. (1995). *Selected Annotated Bibliography on Language Minority Assessment.* Cambridge, MA: FairTest.

Bredekamp, S., & Copple, C. (Eds.). (1997). *Developmentally Appropriate Practice in Early Childhood Programs* (rev. ed.). Washington, DC: National Association for the Education of Young Children.

Medina, N., and Neill, D. M. (1990). *Fallout From the Testing Explosion: How 100 Million Standardized Exams Undermine Equity and Excellence in America's Public Schools* (third ed.). Cambridge, MA: FairTest

National Center on Educational Outcomes and Parents Engaged in Educational Reform (PEER) Project. (1997). *Understanding Educational Assessment and Accountability.* Boston: PEER.

National Forum on Assessment. (1995). *Principles and Indicators for Student Assessment Systems.* Cambridge, MA: FairTest.

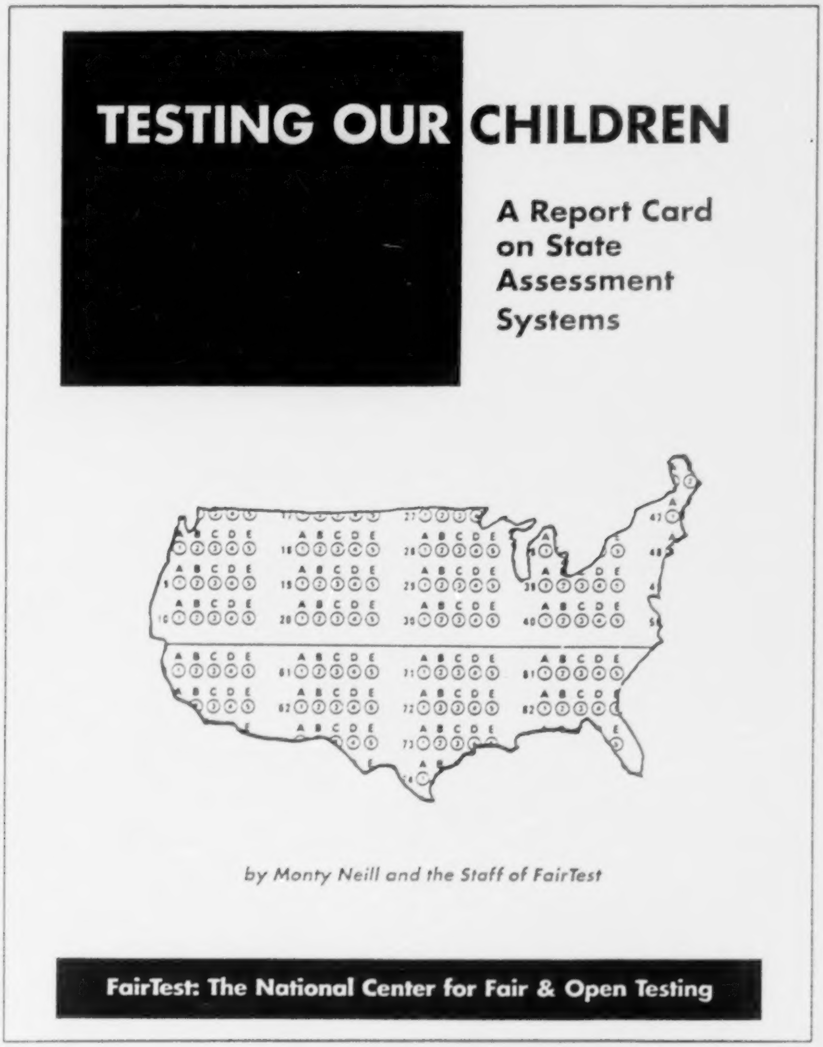
Neill, D. M. (1997). "Transforming Student Assessment." *Phi Delta Kappan*, September 1997.

Neill, D. M., & Medina, N. J. (1989). "Standardized testing: Harmful to educational health." *Phi Delta Kappan, 70*, 688-697.

Neill, M., Bursh, P., Schaeffer, R., Thall, C., Yohe, M., & Zappardino, P. (1995). *Implementing Performance Assessment: A Guide to Classroom, School and System Reform.* Cambridge, MA: FairTest.

Rivera, C., Vincent, C., Hafner, A., & LaCelle-Peterson. (n.d., 1995-96). *Statewide Assessment Programs: Policies and Practices for the Inclusion of Limited English Proficient Students.* Arlington, VA: The George Washington University Evaluation Assistance Center East.

Smith, M. L. (1996). *The Politics of Assessment: A View from the Political Culture of Arizona.* CSE Technical Report 420). Los Angeles: University of California at Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).



TESTING OUR CHILDREN

A Report Card
on State
Assessment
Systems

by Monty Neill and the Staff of FairTest

FairTest: The National Center for Fair & Open Testing

Testing Our Children: A Report Card on State Assessment Systems $______
(FairTest, 1997: 250 pp.) Individual copies for $30; 5 for $125.

Testing Our Children: Executive Summary $______
(FairTest, 1997: 40 pp.) Individual copies for $10; 5 for $40.

Principles and Indicators for Student Assessment Systems $______
Individual copies for $10; 10 for $80; 50 for $350; 100 for $600.

FairTest Examiner, published quarterly; $45/year (institutions), $30/year (individuals) $______

Name ______________________ **TOTAL** $______

Address ______________________

City ____________ State ________ ZIP ________

Make checks payable to: FairTest, 342 Broadway, Cambridge, MA 02139